INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



TEORIA DE NUDOS

Agosto 1979

200 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 35
Agosto 1979

INVESTIGACION Y
CIENCIA
Edición en español de **SCIENTIFIC AMERICAN**

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

Ramón Fontarnau y Frances E. Lynd: *La fertilización humana externa;* Cristóbal Mezquita: *El factor de crecimiento nervioso;* Antonio Creus: *Control automático por inteligencia distribuida;* Luis Bou: *Teoría de nudos;* Pedro Pascual: *Neutrones ultrafríos;* Carlos Alonso del Real: *Una mina de sílex neolítica;* Miquel Gich: *La historia del Atlántico;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio.*

Ciencia y sociedad:

Pedro M. Echenique

Libros:

A. Prevosti, J. de Porta y J. Lafuente



LA PORTADA

La ilustración de la portada muestra un trozo de cordel retorcido y entrelazado, cuyos extremos se han unido a fin de obtener un modelo material de los objetos matemáticos abstractos llamados nudos. En matemática, los nudos son por definición curvas unidimensionales trazadas en el espacio tridimensional que comienzan y terminan en un mismo punto, y que no se cortan a sí mismas (véase "Teoría de nudos", por Lee Neuwirth, en este mismo número). La teoría de nudos, que se propone distinguir entre sí las curvas anudadas, es parte de la topología, rama de la geometría que estudia las propiedades de los espacios que permanecen invariables al deformarlos. Por ejemplo, el modelo de la portada puede deformarse hasta una simple vuelta. A pesar de su aspecto, esta figura está desanudada.

Suscripciones:
Prensa Científica, S.A.
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2200 pesetas
Extranjero:
Un año (12 números): 36 U.S.$
Ejemplar atrasado ordinario:
225 pesetas
Ejemplar atrasado extraordinario:
340 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49; Barcelona-10

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239; Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137; Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 3.ª; Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Diseño de la portada de George V. Kelvin

Página	Fuente
7-10	Ilil Arbel
11	R. G. Edwards Universidad de Cambridge
12	R. G. Edwards, Universidad de Cambridge (*arriba*); Davis Division, Department of Embryology, Carnegie Institution of Washington (*centro*); Ilil Arbel (*abajo*)
13	Ilil Arbel
14	Davis Division, Department of Embryology, Carnegie Institution of Washington
15	Ilil Arbel
16	Dominick P. Purpura, Albert Einstein College of Medicine, New York
19	J. S. Chen, Laboratorio de Biología Celular de Roma
20	Rita Levi-Montalcini, Laboratorio de Biología Celular de Roma
21	Carol Donner
22	Rita Levi-Montalcini, Laboratorio de Biología Celular de Roma
23	Albert Miller
24	Albert Miller (*arriba*); Rita Levi-Montalcini, Laboratorio de Biología Celular de Roma (*abajo*)
25-26	Carol Donner
27	Luigi Aloe, Laboratorio de Biología Celular de Roma
31	Intel Corporation
32-42	Gabor Kiss
53-66	George V. Kelvin
68-70	Tom Prentiss
71	Peter W. Bosch, Servicio Geológico Holandés, Heerlen
72-74	Tom Prentiss
77	Instituto Max-von de Laue-Paul Langevin
78-89	Allen Beechel
93-99	Andrew Tomko
100-101	Albert Miller
102-103	Andrew Tomko
104	Albert Miller
107-114	Carlos Asensio, Fernando Baquero y Miguel Alonso
116-120	Ilil Arbel
123	Macbeth (*arriba*), Michael Goodman (*abajo*)
124-126	Michael Goodman
127	Michael Goodman
128	Ralph C. Eagle

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona
Printed in Spain - Impreso en España

Los autores

CLIFFORD GROBSTEIN ("La fertilización humana externa") es profesor de ciencias biológicas y política pública en el Program in Science, Technology and Public Affairs de la Universidad de California en San Diego. Obtuvo su Ph.D. en zoología en la Universidad de California en Los Angeles y pasó a enseñar en el State College de Oregón. En 1946, después de tres años en la Fuerza Aérea como fisiólogo de la aviación, se unió al equipo del National Cancer Institute y se dedicó a la investigación en el campo de la biología del desarrollo durante los siguientes 10 años. Unióse luego al cuerpo docente de la Universidad de Stanford. Después de nueve años como profesor de biología en Stanford, Grobstein se trasladó a San Diego, donde desde 1967 a 1973 fue decano de la Facultad de Medicina.

RITA LEVI-MONTALCINI y PIETRO CALISSANO ("El factor de crecimiento nervioso") son neurobiólogos del laboratorio de Biología Celular de Roma, perteneciente al Consejo Nacional de Investigación de Italia. Levi-Montalcini recibió el grado de doctor en medicina por la Universidad de Turín en 1936. Como consecuencia de la invasión de Italia por los alemanes, en 1943 huyó a Florencia donde vivió clandestinamente hasta la liberación de Italia en 1945. En 1969 le fue ofrecida la dirección del nuevo laboratorio de Biología Celular. Calissano es jefe del grupo de investigación de neurobiología del Laboratorio de Biología Celular. Graduado en 1964 en la Facultad de Medicina de la Universidad de Génova, trabajó al principio en enzimología. En 1966, se unió al grupo Levi-Montalcini de la Universidad de Washington. A su vuelta a Italia, en 1968, se incorporó como miembro permanente a la unidad de neurobiología de Roma, trasladándose al Laboratorio de Biología Celular en 1974.

STEPHEN KAHNE, IRVING LEFKOWITZ y CHARLES ROSE ("Control automático por inteligencia distribuida") son ingenieros de sistemas en el Case Institute of Technology de la Case Western Reserve University. Stephen Kahne es profesor y presidente del departamento de ingeniería de sistemas. Se formó en las universidades de Cornell e Illinois, obteniendo su doctorado en ingeniería eléctrica en 1963. Trabajó en las Fuerzas Aéreas y en la Universidad de Minnesota. Entró en Case en 1976. Lefkowitz enseña ingeniería de sistemas. Se graduó en la Cooper Union School of Engineering y trabajó durante 10 años en la industria. Se doctoró en Case. Desde 1957 Lefkowitz ha dirigido el programa de investigación de dicho centro superior sobre el control de sistemas complejos. Rose es profesor asociado en ingeniería de computadores. Estudió en la Universidad de Duke, y, tras 4 años como ingeniero jefe de la Texas Instruments Corporation, se doctoró en ciencia de computadores y de información en 1970, en Case.

LEE NEUWIRTH ("Teoría de nudos") dirige la Communications Research Division del Institute for Defense Analysis, en Princeton. Recibió su primer título de ingeniería química de la Universidad de Princeton, su licenciatura en matemáticas aplicada de la Universidad de Columbia y su doctorado en matemáticas, de la Universidad de Princeton, en 1959. Ya doctorado permaneció algún tiempo en la Universidad de California en Berkeley, y en 1961 ingresó en el Institute for Defense Analysis. Fue nombrado director delegado en 1965, y director, en 1977.

PETER W. BOSCH ("Una mina de sílex neolítica") presta sus servicios de cartógrafo en el departamento de mapas geológicos en Servicio Geológico Holandés en Heerlen. Es el responsable de la cartografía de la provincia de Limburgo, en el sur del país. Y se ha especializado en la geohistoria del río Maas (Mosa). Durante nueve años trabajó como voluntario en la excavación de las minas prehistóricas de sílex en Rijckholt, descritas en este artículo.

R. GOLUB, W. MAMPE, J. M. PENDLEBURY y P. AGERON ("Neutrones ultrafríos") son físicos nucleares que han trabajado juntos durante varios años en el desarrollo e instalación de aparatos experimentales en los que se usan neutrones ultrafríos en el Instituto Laue-Langevin (ILL) de Grenoble. Golub se doctoró en el Instituto de Tecnología de Massachusetts en 1967. Tras dos años en Brandeis obtuvo un contrato con la Universidad de Sussex. Mampe consiguió el doctorado en 1971 en la Universidad Técnica de Munich, trasladándose luego al ILL para investigar sobre espectroscopía nuclear. Pendlebury es lector en física nuclear en la Universidad de Sussex. Procede de Cambridge, donde estudió matemáticas, física y química y se doctoró en 1963 en espectroscopía de haces atómicos. Ageron se formó en la Ecole Polytechnique de París. En 1960 entró a formar parte de la Comisión de Energía Atómica Francesa para trabajar en el campo de la investigación de reactores, en particular en el diseño de reactores de alto flujo.

JOHN G. SCLATER y CHRISTOPHER TAPSCOTT ("La historia del Atlántico") son geofísicos cuyo interés se centra en la configuración que presentaban en el pasado las placas tectónicas terrestres. Sclater es profesor de ciencias planetarias y de la tierra en el programa oceanográfico que llevan en común el Instituto de Tecnología de Massachusetts y la Woods Hole Oceanographic Institution. Se licenció en física por la Universidad de Edimburgo trabajando a continuación en geofísica en la Universidad de Cambridge donde se doctoró en 1966. Hasta 1972 investigó en la Scripps Institution of Oceanography y pasó luego al MIT. Sclater ha recibido el Rosensteil Award de Oceanografía correspondiente a 1979. Tapscott es geofísico e investiga para el Exxon Production Research Company, de Houston (Texas). Se licenció en el Swarthmore College en 1972, trabajando posteriormente en física nuclear en la Universidad de Princeton. Al año siguiente obtuvo permiso para seguir estudios de geología marina y geofísica. A finales de 1974 entró a formar parte del programa común de oceanografía realizando su investigación doctoral en el laboratorio de Sclater.

CARLOS ASENSIO y FERNANDO BAQUERO ("Las microcinas") trabajan en Madrid, el primero en el Instituto de Enzimología y Patología Molecular del Consejo Superior de Investigaciones Científicas (CSIC), del que es vicedirector, y en el Centro Especial Ramón y Cajal el doctor Baquero, en donde dirige el Servicio de Bacteriología. Asensio es profesor de investigación y consejero de número del CSIC y preside la Sociedad Española de Bioquímica. Ha trabajado en los Estados Unidos (1959-61) con el profesor B. L. Horecker en la Universidad de Nueva York y con el profesor C. B. van Niel en la de Stanford. Baquero, que en su formación profesional destaca sus antecedentes familiares y su relación constante con el Instituto Pasteur, ha dirigido con anterioridad el Servicio de Microbiología de la Clínica Infantil y de la C. S. "La Paz". Ha sido Presidente de la Sociedad Española de Microbiología, y actualmente preside la Comisión Nacional de la Especialidad de Microbiología.

Hace...

José M.ª López Piñero

...cuatrocientos años

Apareció la cuarta edición de *Morborum internorum curatio* de Miguel Juan Pascual, uno de los mejores compendios de medicina práctica publicados en el siglo XVI. Fue muy apreciado en su tiempo, como lo demuestra el hecho de que alcanzara once ediciones, la primera de las cuales se imprimió en Valencia el año 1555 y la última, en Lyon, más de un siglo después (1664). A partir precisamente de la cuarta, figuraron en todas ellas unos comentarios a su texto redactados por Pedro Pablo Pereda.

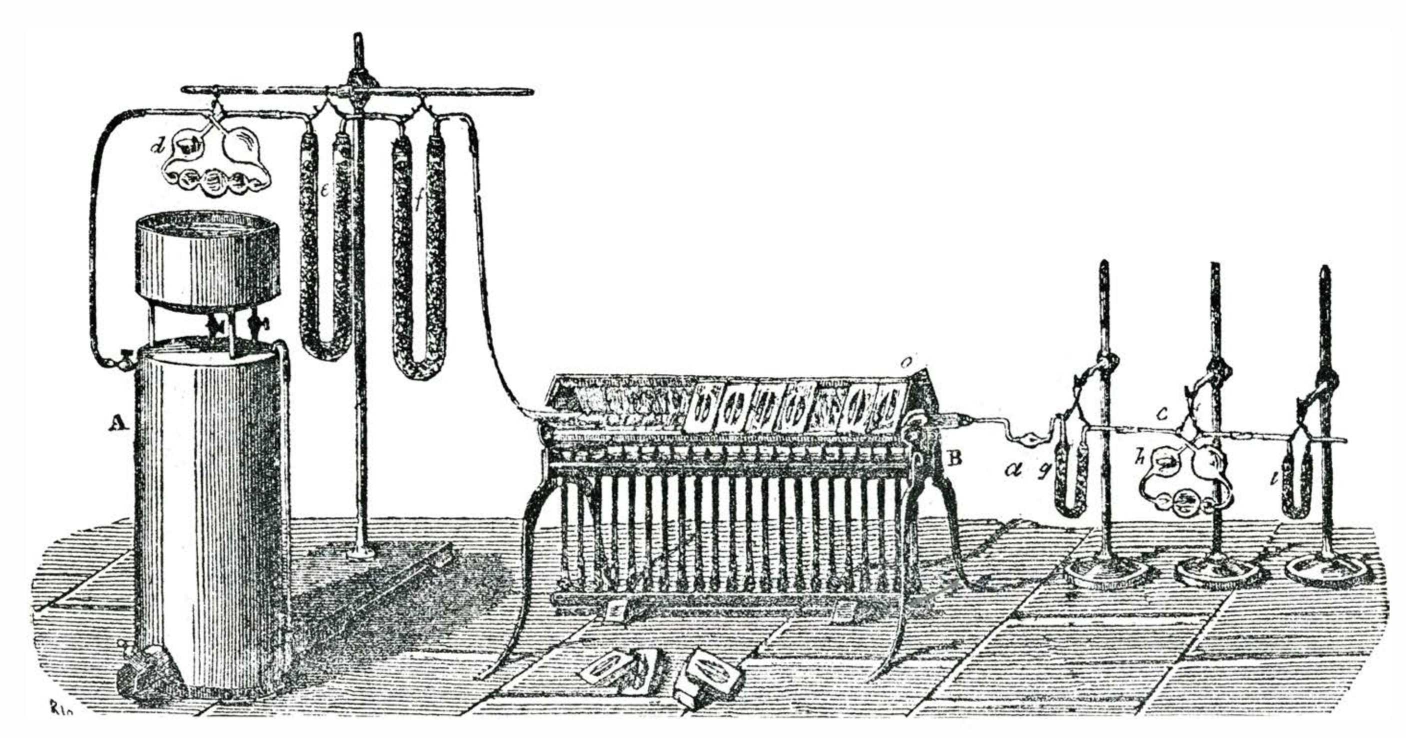
Miguel Juan Pascual se formó como médico en las Universidades de Alcalá y Montpellier y fue luego, durante dos décadas, profesor de medicina en la de Valencia. Puede ser considerado como un seguidor del llamado galenismo "hipocratista", ya que, sin romper con la patología tradicional, pasó a un primer plano la observación clínica y el estudio ambientalista de las enfermedades. Su compendio incluye una primera parte dedicada a afecciones localizadas, en el orden habitual "a capite ad calces", y otra segunda sobre las fiebres. Varios capítulos contienen excelentes descripciones clínicas, entre ellos el consagrado a la sífilis, que fue reimpreso en la importante colección de textos venereológicos de Luigi Luvigini en 1567 y 1728. El enfoque ambientalista se refleja en el estudio de las "enfermedades dominantes" en Valencia en años determinados, así como en una *Medica disputatio* que figura como apéndice de la obra. Se trata de un temprano texto sobre un problema concreto de higiene pública, que examina si la fetidez resultante de la maceración del cáñamo en balsar puede contaminar el aire y ser causa de epidemias.

El profesor valenciano había publicado anteriormente una traducción castellana (1537) del tratado de cirugía de Giovanni da Vigo, que completó con numerosas "adiciones marginales", las más interesantes de las cuales son las de tema anatómico.

...trescientos años

Publicó uno de sus *Lunarios* Carlos de Sigüenza y Góngora, cabeza de la renovación de las ciencias fisicomatemáticas en el México de finales del siglo XVII. Cosmógrafo real y catedrático de matemáticas en la Universidad de México, además de un notable cultivador de la astronomía, la cartografía y las matemáticas, fue un erudito original, poseedor de valiosas colecciones, entre las que se encontraba el códice náhuatl hoy llamado *Jeroglífico de Sigüenza.*

En sus *Lunarios* incluyó cuidadosas observaciones astronómicas y demostró que carecía de convicción en la astrología. Con motivo del cometa de 1680, mantuvo una polémica con varios autores que consideraban los cometas como portentos o alteraciones del orden natural. Midió las paralajes del citado cometa de 1680 para probar su ubicación supralunar, como había hecho un siglo antes Jerónimo Muñoz. Observó asimismo, con telescopio y cuadrante, el eclipse solar del 21 de agosto de 1691 y determinó la longitud de la ciudad de México, utilizando las tablas astronómicas de Tycho Brahe. Defendió un sistema cosmológico similar al de este último autor, utilizando elementos aislados del copernicanismo. Su principal publicación en este campo fue el libro titulado *Libra astronómica y philosophica* (1690).



Aparato para el análisis elemental de las sustancias orgánicas según Gabriel de la Puerta. B *y* o, *hornillo de gas donde se coloca el tubo de combustión;* a, *tubo con una esferita para que se condense el agua;* g, *tubo en U con cloruro de calcio en su primer brazo y con piedra pómez impregnada de ácido sulfúrico en el segundo;* h, *tubo de bolas con disolución de potasa cáustica de 45º;* i, *tubo en U con piedra pómez impregnada de disolución de potasa en su primer brazo, y con pedazos de potasa en el segundo:* A, *gasómetro con oxígeno para hacerle pasar por el tubo de combustión al final de la operación;* d, *tubo de bolas con potasa cáustica, para que retenga el ácido carbónico que pueda tener el oxígeno;* e *y* f, *dos tubos grandes en U, con cloruro de calcio el primero y con piedra pómez y ácido sulfúrico el segundo, para privar al oxígeno de la humedad.*

Como matemático, Sigüenza fue un seguidor de la obra de Descartes. En su obra cartográfica, destaca un mapa de la bahía de Pensacola que trazó tras haber participado como cosmógrafo en una expedición que, en 1693, recorrió el golfo de México. Su mapa de Nueva España, publicado en 1675, fue el primero que incluyó todo el virreinato. Hizo también un mapa "de las aguas que vienen a la laguna de Texcoco", al ocuparse, en 1691-92, de los problemas de drenaje de la ciudad de México.

...doscientos años

Francesc Sanponts i Roca recibió el título de doctor en medicina por la Universidad de Huesca. Viajó a continuación por el extranjero, donde perfeccionó su preparación científica, sobre todo en física y matemáticas. A su regreso, fijó su residencia en Barcelona, donde permaneció la mayor parte de su vida, desarrollando una gran actividad en materias científicas y técnicas.

Realizó algunas estimables contribuciones a la medicina, como una memoria sobre el *muguet* infantil, premiada en 1786 por la Academia de Medicina de París. Más importantes y numerosos fueron, sin embargo, sus trabajos relacionados con la mecánica y sus aplicaciones técnicas. Ya en 1784, publicó, en colaboración con Francesc Salvá i Campillo, una monografía sobre una agramadora de cáñamo y lino que ambos habían inventado. Dos años más tarde, ingresó en la Academia de Ciencias y Artes de Barcelona, ante la que leyó diversas memorias en su mayor parte relacionadas con la mecánica. En 1804, a petición de un fabricante de tejidos local, diseñó y dirigió la construcción de una máquina de vapor de veinte caballos, que quedó instalada un año después.

De gran fertilidad fue la labor docente que Sanponts realizó al frente de la Escuela de Mecánica, fundada por la Junta de Comercio de Barcelona en 1806 y que, tras la interrupción de la Guerra de Independencia, volvió a funcionar en 1814.

...cien años

Gabriel de la Puerta Ródenas publicó su *Tratado de Química Orgánica,* edición definitiva de otra obra suya anterior y uno de los más destacados textos sobre la materia aparecidos en la España de la época. Sus dos volúmenes ofrecen una síntesis orientada fundamentalmente al estudio de las aplicaciones farmacéuticas e industriales de las sustancias orgánicas. Incluye también una parte general, en la que destaca el capítulo consagrado a las fermentaciones. La obra concluye con un resumen de "química biológica".

Gabriel de la Puerta estudió en la Universidad de Madrid, donde obtuvo el doctorado en farmacia en 1863. Fue primero catedrático supernumerario de química orgánica en la Facultad de Farmacia de la misma Universidad y, más tarde, pasó a ocupar la cátedra de química inorgánica. Dirigió el laboratorio de análisis químico del Ministerio de Hacienda y participó intensamente en la actividad de varias instituciones científicas, entre ellas las Reales Academias de Ciencias y de Medicina y la Sociedad de Física y Química.

La producción escrita de Gabriel de la Puerta comprende centenar y medio de trabajos científicos. Además de la obra citada, sobresale por su importancia un *Tratado de Química Inorgánica* (1896-97), asimismo en dos volúmenes, que publicó al final de su vida.

...cincuenta años

Charles Upson Clark redescubrió dos fuentes de excepcional importancia para la historia de la actividad científica en la América española durante el Renacimiento.

La primera de ellas es el *Compendio y Descripción de las Indias Occidentales* de Antonio Vázquez Espinosa, del que halló, en la Biblioteca Vaticana, un ejemplar integrado en parte por las pruebas de una impresión interrumpida. Su contenido es fundamentalmente geográfico y corresponde al viaje que realizó su autor por México, Centroamérica y Perú en la segunda década del siglo XVII. La descripción es tan detallada –afirma Clark– "que puede servir como autoridad, aun en cuestión de fronteras históricas". Incluye también descripciones de plantas y animales, con numerosos sinónimos en diferentes lenguas americanas, e informa sobre las minas de Potosí, Huencavélica y Oruro.

La segunda fuente redescubierta por Clark –esta vez en colaboración con Lynn Thorndike– fue el *Libellus de medicinalibus indorum herbis,* único texto de botánica médica azteca que ha llegado hasta nosotros. Sus autores fueron Martín de la Cruz y Juan Badiano, médicos indios mexicanos formados en el Colegio Imperial de Santa Cruz de Tlatelolco. Al parecer, el primero redactó el texto en náhuatl y el segundo lo tradujo al latín. Incluye trece capítulos en los que se agrupan las plantas empleadas en terapéutica. Su contenido corresponde plenamente a la cultura india mexicana, aunque sus autores estaban familiarizados con la *Historia natural* de Plinio.

La fertilización humana externa

El procedimiento por el que un nuevo ser humano puede concebirse en una cápsula de vidrio ha planteado serios problemas científicos y éticos. La cuestión ha de debatirse con un cabal conocimiento del desarrollo humano

Clifford Grobstein

El 25 de julio del año pasado venía al mundo Louise Joy Brown (con un peso al nacer de 2,6 kilogramos), hija del matrimonio Gilbert John Brown de Bristol, Inglaterra. Su nacimiento en el Oldham General Hospital cerca de Manchester se consideró en todo el mundo como el comienzo de una nueva era en el campo de la reproducción humana. Constituía el primer caso registrado en que se había extraído un ovocito humano de un ovario de mujer, fertilizado externamente e implantado luego en el útero de aquélla, desarrollándose a término. El éxito del procedimiento abrió la puerta a la esperanza de aquellas parejas que sufren de diversas formas de esterilidad para poder parir sus propios hijos. Cualquier beneficio práctico en este punto de la investigación biomédica suele recibirse como un motivo de satisfacción general. Pero este avance en particular ha suscitado reacciones contrapuestas, y, en algunos sectores, ha sido causa de una seria preocupación.

La fertilización humana externa, que comprende la manipulación de laboratorio de células con potencial para convertirse en ser humano plantea problemas que enlazan con los relativos al aborto y a los experimentos que involucran a los seres humanos; uno y otro han venido siendo objeto de controversia en Occidente y, en particular, en los Estados Unidos durante varios años. Cuando nació en Inglaterra Louise Joy Brown, hacía unos cinco años que la investigación sobre la fertilización externa había sido suspendida en los Estados Unidos, durante los cuales el Congreso primero, una comisión nacional sobre experimentación en humanos después y finalmente una nueva Ethics Advisory Board acordaron aconsejar al Secretary of Health, Education, and Welfare, considerase la idoneidad de diversas formas de investigación "fetal". En marzo de este año la comisión concluyó que la investigación sobre la fertilización externa e implantación subsiguiente "sería aceptable desde un punto de vista ético". Remitía a los National Institutes of Health, para su reconsideración a la luz de diversas reservas expresadas, una proposición que llevaba mucho tiempo de existencia en torno a la investigación de la fertilización humana (no respecto a la implantación) in vitro de Pierre Soupart, de la Facultad de Medicina de la Universidad de Vanderbilt. Se sabe que varios grupos clínicos se están preparando para proporcionar fertilización externa e implantación a mujeres estériles.

El debate sobre el procedimiento habrá de proseguir. El nacimiento de un ser humano concebido en una cápsula de vidrio es suficiente novedad, con implicaciones tanto sociales y éticas como biológicas, como para plantear cuestiones que precisan ser afrontadas por la comunidad biomédica y el público. Un haz de preguntas giran en torno a si la fertilización externa ha sido o no ensayada adecuadamente y demostrado ser lo bastante segura para justificar su aplicación como una forma de terapia para la infertilidad. Aun cuando el procedimiento se aceptase como válido con óptimos resultados, ¿representa cruzar una especie de Rubicón biológico y ético hacia un territorio desconocido? ¿Qué posibilidades futuras abre en la manipulación del desarrollo humano por parte del hombre? ¿Son éstas deseables? Este artículo traza las líneas generales de la base científica para las esperanzas y temores que acompañan al nuevo gran avance y considera alguno de los problemas de principio que plantea.

La fertilización de un ovocito humano in vitro parece haberse conseguido por primera vez en Inglaterra, en 1969, por los fisiólogos R. G. Edwards y B. D. Bavister y el ginecólogo P. C. Steptoe. Edwards y Steptoe se dedicaron a desarrollar la técnica y aplicarla para conseguir el nacimiento en julio del año pasado y luego un segundo nacimiento con éxito en enero último. La investigación sobre la fertilización externa en especies no humanas, sobre todo animales marinos, se remonta, sin embargo, a 1893. Hace dos décadas se presentaron pruebas convincentes de fertilización externa de óvulos de conejo colmada por el éxito. Desde entonces, el proceso se ha establecido en media docena de otras especies de mamíferos, proporcionando considerable información que sirvió para desarrollar la técnica en humanos. Durante los últimos 25 años, la profundización en el conocimiento de los complejos procesos mediante los cuales se llevan a cabo normalmente la fertilización interna y el desarrollo en los mamíferos, y específicamente en los seres humanos, hizo posible la fertilización externa.

De particular importancia resultó un conocimiento detallado de los procesos hormonales interrelacionados que controlan el ciclo reproductor femenino. Los principales elementos anatómicos implicados en el ciclo son el hipotálamo situado cerca de la base del cerebro, la glándula hipófisis justo debajo, los ovarios, el oviducto, el útero y la vagina. En cada ciclo reproductor normal de la mujer, madura un ovocito, es expulsado del ovario y pasa hacia el extremo dilatado del oviducto, por el que se desplaza hasta el útero. En ausencia de inseminación, el ovocito se desintegra y el ciclo se repite a intervalos más o menos mensuales. Si ha habido inseminación en el momento oportuno del ciclo, el ovocito puede encontrar espermatozoides en el oviducto y puede ser fertilizado. El óvulo fertilizado se desplaza luego por el oviducto, al tiempo que se va dividiendo. El huevo alcanza normalmente el útero unos tres días más tarde en forma de mórula, un grupo esférico de ocho o más células. La división de las células continúa en el útero produciendo un blastocisto, una bola vacía con un grupo de células, la masa celular in-

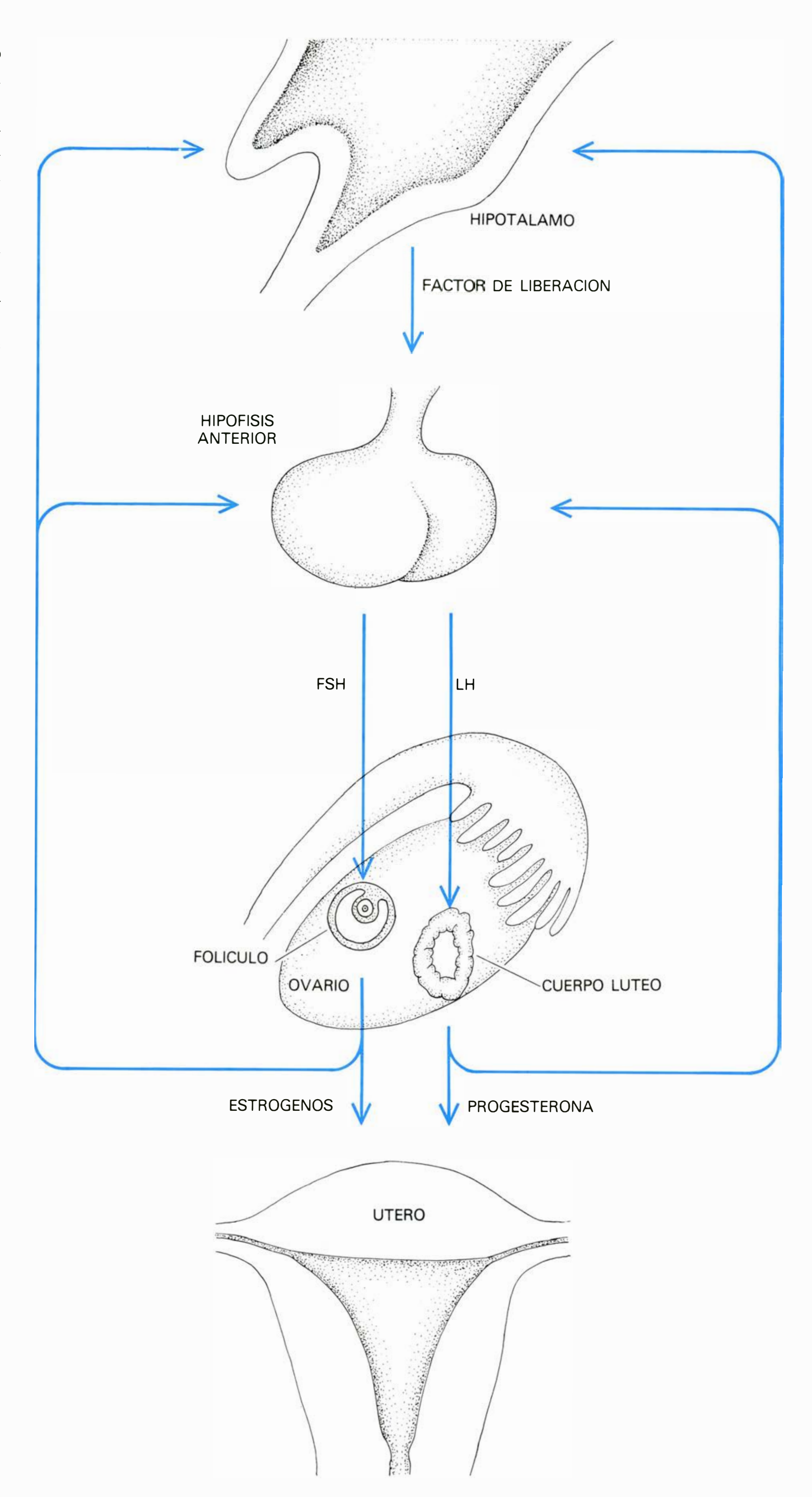
terna, en uno de los polos. El blastocisto se adhiere al endometrio, el forro del útero; un complejo proceso de implantación empotra el blastocisto en la pared uterina, donde prosigue el desarrollo. La masa celular interna se convierte en el embrión; las células externas del blastocisto, junto con las células uterinas vecinas y los vasos sanguíneos, forman la placenta, el órgano que mantiene el intercambio de materiales entre la futura circulación embrionaria y la circulación materna. Con la implantación, puede decirse que ha empezado el embarazo.

Lo que organiza estos acontecimientos es la interacción de hormonas producidas en el hipotálamo, la hipófisis, los ovarios y la placenta. El único ovocito expulsado cada mes de uno u otro de los ovarios madura en un folículo constituido por muchas células. El folículo aumenta de tamaño y en número de células durante la primera parte del ciclo bajo la influencia de la hormona estimulante de los folículos (FSH, del inglés "follicle-stimulating hormone") segregada por la hipófisis. Al crecer el folículo, sus células segregan las hormonas estrógenas, siendo la más importante de entre ellas el estradiol, que actúa tanto sobre la hipófisis como sobre el útero. La retroalimentación de la hipófisis reduce la producción de FSH; en el útero, estimula el crecimiento de la pared y glándulas uterinas.

Al acercarse el folículo a su máximo tamaño, disminuye la producción de FSH, haciéndose predominante una segunda hormona hipofisaria, la hormona luteinizante (LH, del inglés "luteinizing hormone").

La LH completa la maduración del ovocito y provoca la ruptura del folículo, descargando el ovocito al oviducto. Asimismo es la causa de que las células del folículo residual se transformen en el cuerpo lúteo, que segrega una segunda hormona ovárica, la progesterona, la cual actúa sobre la pared uterina para inducir un posterior desarrollo glandular y sensibilizar la reacción del útero al contacto con las células del blastocisto. En ausencia de embarazo (si un blastocisto no llega al útero) empieza a crecer un nuevo grupo de folículos ováricos, y el cuerpo lúteo derivado del folículo ovulado empieza a declinar. El cambio resultante en el equilibrio estradiol-progesterona provoca a su vez cambios degenerativos en la pared uterina y empieza la menstruación.

Pero si un óvulo ha sido fertilizado y llega un blastocisto al útero, las células externas del blastocisto segregan gonadotropina coriónica humana (HCG, del



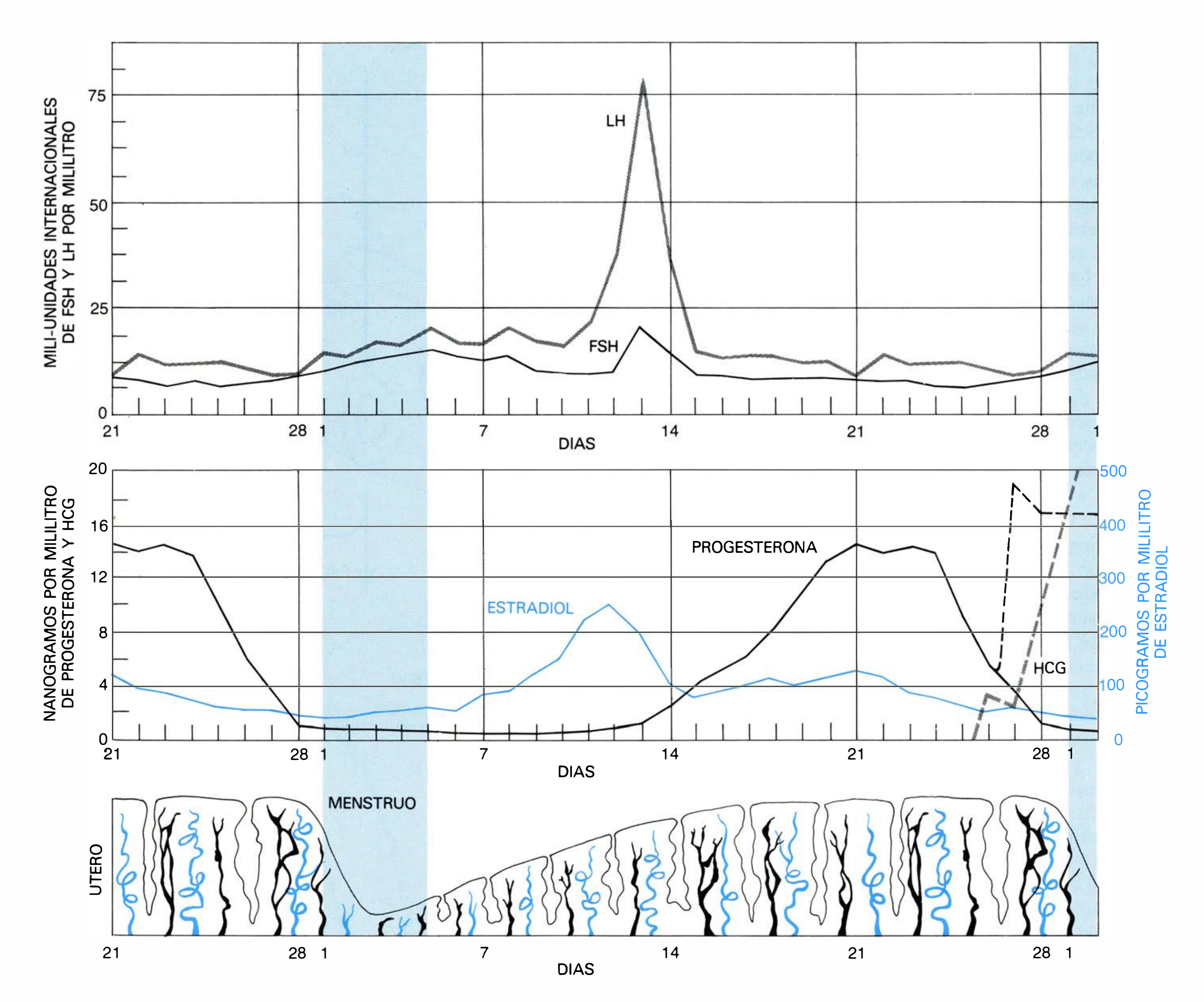
ACTIVIDAD HORMONAL INTERRELACIONADA que controla el ciclo reproductor femenino. Un factor de liberación segregado por el hipotálamo, situado en la base del cerebro, estimula la hipófisis anterior adyacente para que segregue la hormona estimulante de los folículos (FSH) y la hormona luteinizante (LH) vertiéndolas a la circulación. La FSH da lugar a que uno o más folículos del ovario crezcan y produzcan estrógenos, sobre todo estradiol; la LH pone en funcionamiento la ovulación y la transformación del folículo en un cuerpo lúteo, que segrega progesterona. Los estrógenos y la progesterona circulan hacia el útero y preparan la pared uterina para la implantación del huevo; las hormonas ováricas inhiben o estimulan, por retroalimentación, las secreciones del hipotálamo y de la hipófisis.

inglés "human chorionic gonadotropin"), una hormona que tiene el mismo efecto que la LH hipofisiaria. Con ello, el cuerpo lúteo ovárico se mantiene y continúa segregando progesterona; no crecen nuevos folículos en el ovario y la menstruación no tiene lugar. Alrededor de la quinta semana la placenta toma el mando, segregando progesterona para mantener el embarazo.

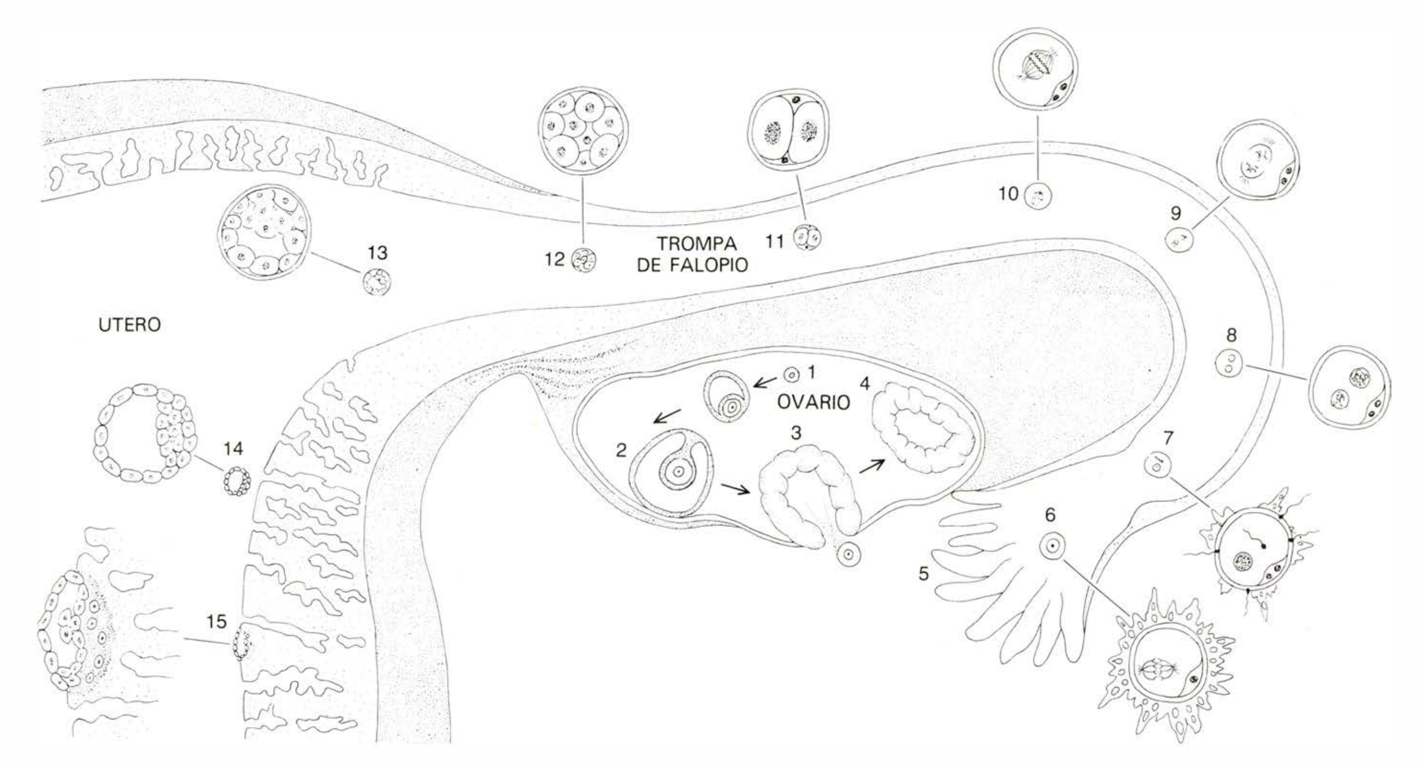
Estos efectos hormonales que controlan el ciclo menstrual tienen diversas implicaciones para el éxito de la fertilización externa. La administración de FSH inmediatamente después de la menstruación puede conducir a la maduración de varios folículos en vez de uno solo, dando "superovulación" y nacimientos múltiples. El aumento de frecuencia en la producción de quintillizos humanos, e incluso sextillizos en los últimos años, refleja la administración más frecuente de FSH para estimular la producción de ovocitos en el tratamiento de la infertilidad femenina debida a una producción reducida de óvulos.

La técnica puede aumentar también el suministro de ovocitos para la fertilización in vitro. Los ovocitos se obtienen del ovario mediante laparoscopia, procedimiento mínimamente traumático desarrollado con finalidad diagnóstica. Un diminuto telescopio y un iluminador de fibra óptica se introducen a través de una incisión muy pequeña practicada en la región del ombligo; pueden examinarse el ovario y cualquier folículo maduro, pudiéndose extraer por aspiración los ovocitos de los folículos. Hay considerable número de datos que evidencian que solamente alrededor de uno de cada cuatro óvulos formados y fertilizados en una hembra sana da descendencia normal a término, y por ende, en teoría, es una ventaja obtener múltiples ovocitos para estudios de fertilización in vitro.

Algunas pacientes han sido por ello tratadas con FSH antes de la laparoscopia para aumentar el número de ovocitos obtenidos durante una única intervención. Puede también ser deseable tratar a la paciente con LH para asegurar la completa madurez de los ovocitos. Durante el desarrollo y prueba del procedimiento, la inducción de superovulación podría tener además la ventaja de pro-



NIVELES HORMONALES EN SANGRE que gobiernan el ciclo reproductor representados junto con cambios en el espesor (*abajo*) del forro del útero y en la proliferación de sus glándulas y vasos sanguíneos. Un lento y precoz aumento del nivel de FSH hace que los ovarios segreguen estradiol, que provoca el engrosamiento y vascularización del endometrio. Un máximo de LH lleva a la ovulación y a la formación de un cuerpo lúteo que segrega progesterona, preparando aún más el endometrio para la implantación. En ausencia de embarazo, la secreción de progesterona decrece y el endometrio se desprende. Sin embargo, si un óvulo es fertilizado y llega a la fase del blastocisto, primero el blastocisto y luego la placenta segregan HCG (del inglés "human chorionic gonadotropin"), una hormona similar a la LH (del inglés "luteinizing y hormone"). La HCG (*curva gris discontinua*) hace que el cuerpo lúteo continúe segregando progesterona (*curva negra discontinua*) según los datos aportados por Christopher A. Adejuwon.



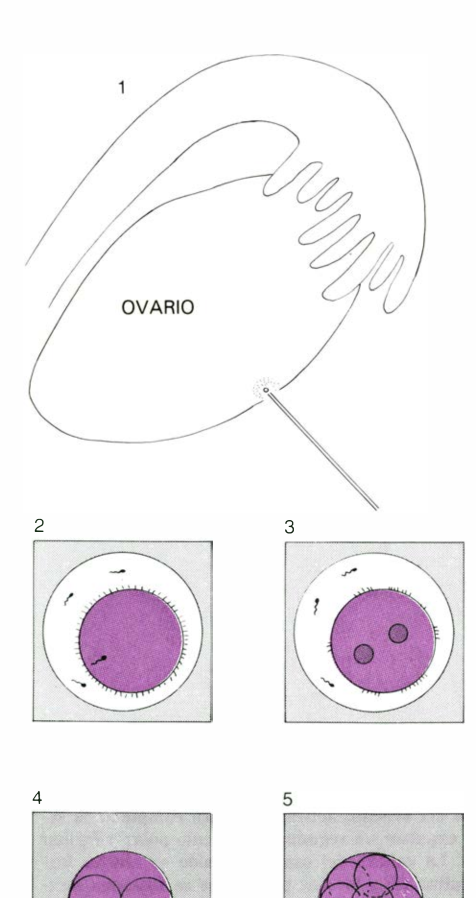
FERTILIZACION INTERNA de un óvulo humano. Es un acontecimiento dentro de un complejo proceso que empieza cuando un folículo primario del ovario (*1*) se desarrolla para convertirse en un folículo maduro (*2*), que se rompe liberando el ovocito (*3*). El folículo se convierte en un cuerpo lúteo (*4*). (El desarrollo folicular, para mayor claridad, se representa arbitrariamente en diferentes lugares del ovario.) La célula desfoliculada es arrastrada de la superficie del ovario por el extremo abierto del oviducto, o trompa de Falopio (*5*). El ovocito acaba de completar la primera parte de la meiosis, una división por la cual sus cromosomas son reducidos a la mitad de la dotación normal y es expulsado un corpúsculo polar al borde del ovocito; ha empezado la segunda parte de la meiosis (*6*). Penetra un espermatozoide en el citoplasma del ovocito, activándolo para completar la segunda parte de la meiosis, expulsar un segundo corpúsculo polar y formar el pronúcleo femenino (*7*). La cabeza del espermatozoide se hincha formando un pronúcleo masculino (*8*) y los dos pronúcleos se fusionan mezclando los cromosomas masculinos con los femeninos (*9*). Los cromosomas se duplican y dividen (*10*) y el óvulo fertilizado, o cigoto, experimenta división (*11*). Las sucesivas divisiones mientras el cigoto se desplaza producen una mórula (*12*), un joven blastocisto (*13*) y, después de unos cuatro días y medio, un blastocisto maduro con una masa celular interna, que recibe el nombre de embrioblasto, que da lugar al embrión, y un trofoblasto externo (*14*). Al sexto o séptimo día, el blastocisto se implanta en la pared (*15*).

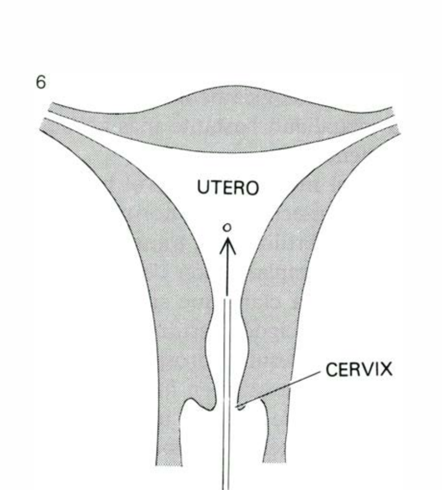
ducir varios ovocitos para estudio, y, sobre todo, para el desarrollo de mejores técnicas destinadas a mantener los ovocitos externamente. El mantenimiento externo no es un problema banal. El óvulo está normalmente en suspensión en los complejos líquidos biológicos presentes en el oviducto y en el útero. No se sabía bien al principio qué constituyentes de estos líquidos eran los esenciales para garantizar el buen estado permanente de los óvulos. Fue precisa la experiencia del cultivo de óvulos de otros mamíferos y de células en general para obtener una base con la que formular los requerimientos de los óvulos humanos.

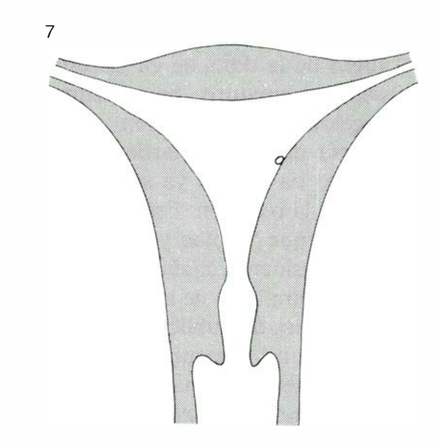
Tenía que abordarse un problema similar respecto al esperma, el otro componente de la fertilización normal. Los espermatozoides se obtienen en gran número del semen eyaculado. Sin embargo, había que superar un importante obstáculo antes de poderse llevar a cabo la fertilización humana externa. En el eyaculado inicial, los espermatozoides no están preparados para penetrar el ovocito. Primero deben experimentar una capacitación, un proceso que normalmente es inducido por ciertas condiciones del tracto genital femenino. Estas condiciones deben simularse externamente para que los espermatozoides se hallen en el estado adecuado para penetrar el ovocito. Teniendo en cuenta los requerimientos tanto del ovocito como de los espermatozoides, había que definir cuidadosamente un medio artificial apropiado para su unión; la idoneidad del medio tenía que ser validada comprobando que un espermatozoide entra en un ovocito intacto, y que tanto el ovocito como el espermatozoide funcionan luego normalmente.

Aunque hay pasos intermedios en la demostración, la prueba final es el normal desarrollo del cigoto u óvulo fertilizado. En el caso de huevos que se desarrollan externamente, por ejemplo los de la estrella de mar en agua salada o los de las ranas en agua dulce, esto no era un gran obstáculo; ahora bien, para los huevos que se desarrollan internamente, el problema se hace difícil. Existen dos opciones: la vuelta del huevo a su ambiente interno normal o el desarrollo de un ambiente externo adecuado. Como procedimiento terapéutico en los seres humanos no hay duda de que la primera opción es biológicamente más simple y en la actualidad bastante más aceptable socialmente.

Dado el minucioso control hormonal de los acontecimientos normales en la ovulación, fertilización, transporte por el oviducto e implantación del huevo, sin embargo, está claro que si la implantación de un cigoto fertilizado externamente ha de resultar próspera, el estado del cigoto debe estar en íntima armonía con el estado de la pared uterina. Si, por ejemplo, una célula reproductora femenina, fertilizada después de su extracción de un folículo, se desarrolla más lentamente fuera que dentro del oviducto, el cigoto podría ser introducido en el útero de la donante en un estadio demasiado prematuro para estimular una reacción de implantación o podría introducirse demasiado tarde debido a que el útero ha pasado ya su momento óptimo para la reacción. Esta era una de las variables que Steptoe y Edwards tenían que considerar cuidadosamente, y los detalles completos de su control aún están pendientes de publicación. La probabilidad de una armonía eficaz puede influenciarse teóricamente con la administración a la madre de hormonas ade-







cuadas, pero Steptoe ha indicado que esto no se practicó en sus casos logrados, debido a que los niveles hormonales no usuales tienen efectos adversos.

La fertilización externa seguida de implantación, según informaron Steptoe y Edwards, está basada en una confluencia de nuevos conocimientos: sobre biología de la reproducción, sobre efectos hormonales, sobre la función celular (y particularmente nuclear) y sobre el mantenimiento de células, órganos e incluso embriones enteros bajo condiciones artificiales. El procedimiento no es un truco aislado concebido ingeniosamente por dos hombres frente a un problema individual. Es más bien una realización específica –una aplicación particular– de nuevos conocimientos acumulados en muchas investigaciones por muchas personas durante muchos años. Si la fertilización humana in vitro fuere regulada o incluso prohibida, no por ello se borrarían las crecientes posibilidades de este procedimiento y su desarrollo posterior.

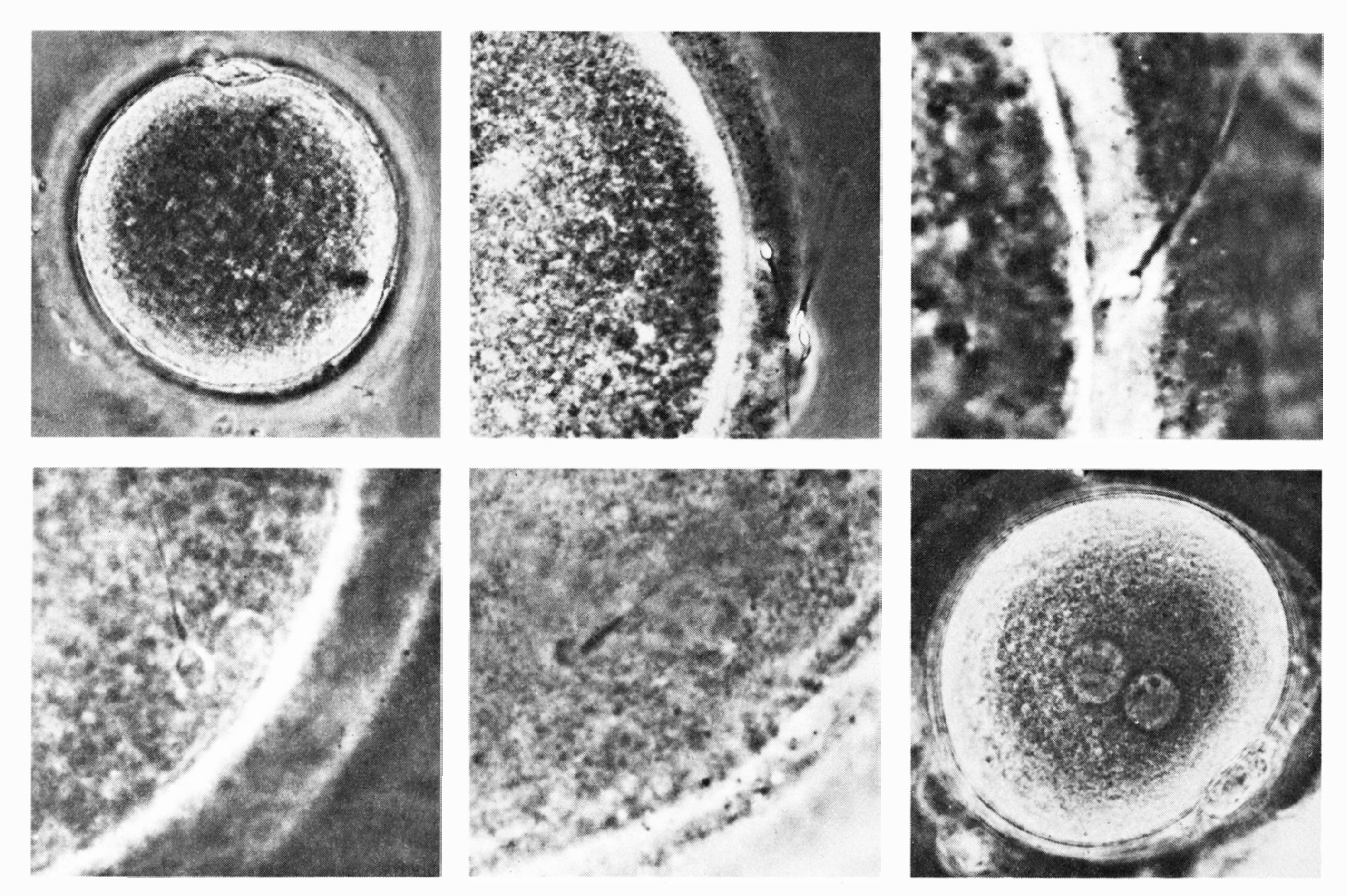
Este es el quid de las cuestiones que se plantean. El Secretario de Health, Education and Welfare, Joseph A. Califano, Jr., citó los temas más importantes en un memorándum dirigido a la comisión ética el pasado septiembre. "Las técnicas de fertilización in vitro y trasplante del embrión ¿pueden dañar al feto resultante y dar lugar a niños anormales? ¿Llevará esta investigación a una reproducción selectiva, a intentos de controlar la estructura genética de la descendencia o al empleo de "padres subrogativos" donde, por ejemplo, las mujeres ricas paguen a mujeres pobres para dar a luz?"

Para considerar la primera de las preguntas del secretario Califano, no hay duda de que, cuando se aplica una nueva técnica a los seres humanos por primera vez, acarrea cierto riesgo. En el caso presente el juicio médico es que hay sólo un ligero riesgo para la madre y un riesgo de magnitud algo menos definida para el feto. La madre está expuesta a los peligros de cualquier tratamiento hormonal, de la laparoscopia y del trasplante del embrión a su útero y a los riesgos normales del embarazo. Los procedimientos hormonales y quirúrgicos se consideran relativamente menos importantes en ese sentido. Quizás el mayor riesgo (sin llegar a ser grande) es el mismo embarazo, riesgo que se supone conocido por la paciente y aceptado voluntariamente.

FERTILIZACION EXTERNA. Procedimiento desarrollado por R. G. Edwards, de la Universidad de Cambridge, y P. C. Steptoe, que no ha sido publicado en detalle; el procedimiento que aquí se esquematiza se basa en registros de charlas y entrevistas dadas por Steptoe. Cuando un aumento del nivel de LH señala una inminente ovulación, se extrae un ovocito del folículo maduro por laparoscopia (*1*), se lava en un medio de cultivo y se coloca en una gotita que contiene espermatozoides del marido (*2*). Se revela la fertilización por la aparición de los dos pronúcleos (*3*). El óvulo fertilizado se transfiere a un medio de cultivo distinto y se observa su división (*4*). O bien en el estadio de ocho células (*5*) o en el de 16, el agregado celular se introduce, por medio de una fina cánula, en el útero (*6*), donde se implanta (*7*).

El riesgo para el feto es, actualmente, más difícil de valorar y debe medirse a la luz del alto riesgo a que cualquier huevo en desarrollo está expuesto en el curso de la fertilización e implantación naturales. Se estima que de cada 100 ovocitos humanos expuestos al esperma, sólo 84 son fertilizados y únicamente 69 se implantan. Se estima que una semana más tarde sólo 42 de los 100 están todavía vivos; de ellos, 37 sobreviven a la sexta semana del embarazo y sólo 31 llegan a término. El proceso normal de la fertilización interna e implantación implica, pues, una pérdida de un 70 por ciento; y de los niños recién nacidos, un pequeño porcentaje presentará anormalidades del desarrollo de diversa consideración. Este alto grado de pérdida embrionaria y anormalidad se cree que se debe, en gran parte, a aberraciones cromosómicas, producidas principalmente durante la formación de los óvulos y de los espermatozoides y, en un grado inferior, durante la propia fertilización. El riesgo natural encierra un complejo transfondo en función del cual se calculan las consecuencias de la fertilización in vitro seguida de implantación.

Los dos nacimientos con éxito registrados no eliminan la preocupación de que el procedimiento puede no ser seguro para el feto, si bien indican que algunos de los peores temores –de niños con grandes anormalidades– no se dan, por lo menos, en todos los casos. El progreso intrauterino de los fetos en desarrollo, según se informó, fue seguido cuidadosamente por medio de diversos procedimientos de monitorización estándar, que no revelaron ningún dato de anormalidad, según cuentan, en ninguna de las fases. Y aún quedan preguntas por contestar. Steptoe dijo el pasado invierno que un total de 79 mujeres habían entrado en tratamiento desde que él y su colega empezaron ensayos clínicos: 68 fueron sometidas a laparoscopia, de las cuales 44 proporcionaron ovocitos en estado correcto de madurez. La fertilización y subsiguiente división del huevo fueron un éxito en 32 casos; cuatro mujeres quedaron embarazadas, de las cuales dos llevaron el feto a término y dos abortaron. Se dispone de poca in-



UNA CELULA REPRODUCTORA FEMENINA HUMANA es fertilizada en una cápsula de vidrio de laboratorio, según micrografías tomadas por Edwards. El ovocito no fertilizado, con su capa glucoproteica protectora, la zona pelúcida, tiene unos 0,2 milímetros de diámetro (*arriba a la izquierda*); el corpúsculo polar, expulsado durante la división meiótica justo antes de la ovulación, está en el borde superior del ovocito. Se muestran segmentos del ovocito, primero con dos espermatozoides penetrando la zona pelúcida (*arriba en el centro*) y luego con un espermatozoide en el espacio perivitelino (*arriba a la derecha*). Puede verse un espermatozoide dentro del citoplasma (*abajo a la izquierda*). La pieza media de un espermatozoide se ve en el citoplasma de un óvulo fijado y teñido después de empezada la fertilización; la cabeza del espermatozoide comienza a hincharse para formar un pronúcleo masculino (*abajo en el centro*). Los pronúcleos masculino y femenino, que se fusionan antes de la primera división, se ven en el cigoto vivo.

formación sobre las causas o consecuencias de los fracasos y de cómo su frecuencia y naturaleza se comparan con el modelo de la reproducción normal. Además, los frutos logrados con el procedimiento están aún en su infancia y es posible que puedan manifestarse todavía algunos efectos a largo plazo.

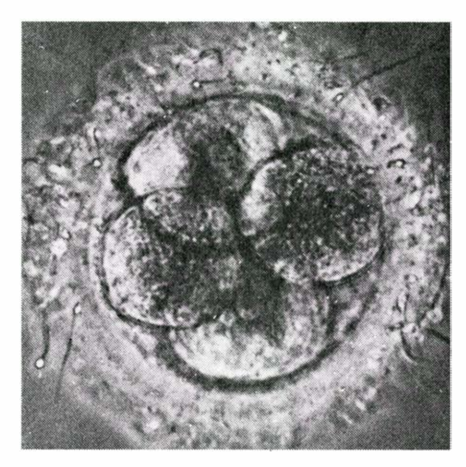
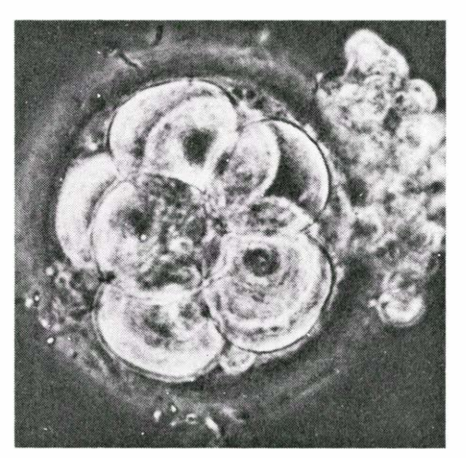
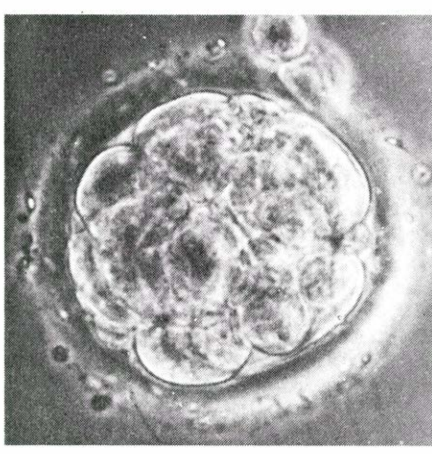
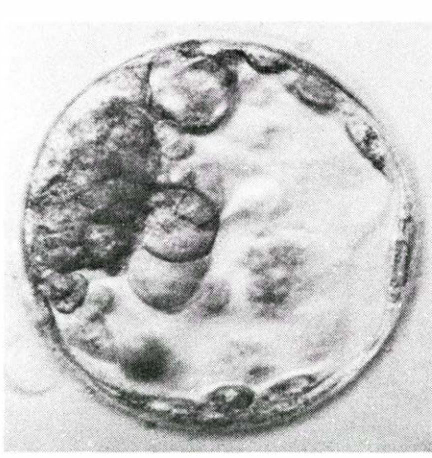
Por último, cuatro embarazos no son suficientes para basar en ellos una conclusión estadística acerca del nivel de riesgo que corre el feto en desarrollo.

Un planteamiento prudente justifica, pues, precaución, como en el caso de cualquier terapia innovadora. Por el momento hay disparidad de criterios entre los expertos. Steptoe y Edwards están claramente convencidos de que los experimentos anteriores tanto con animales como con seres humanos justificaban los ensayos clínicos.

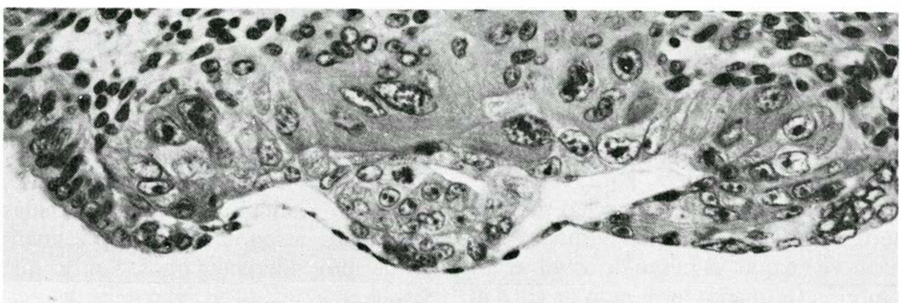
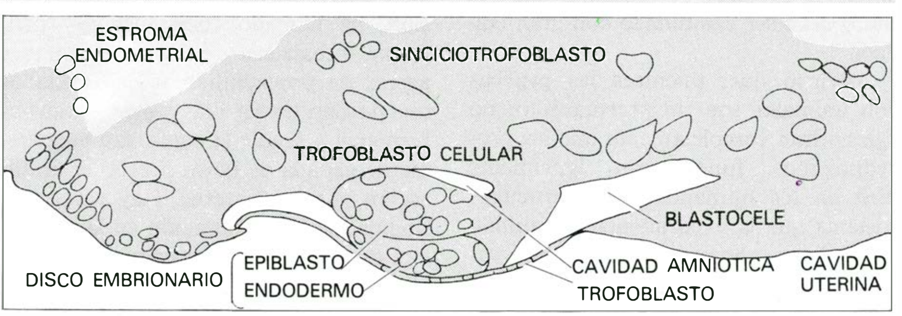
Otros expertos no están aún completamente persuadidos. La solicitud de Soupart era sobre una investigación explícitamente orientada a obtener más información en el proceso de fertilización, considerando la existencia de cierta incertidumbre residual. En la primera reunión de la comisión para considerar la solicitud en septiembre del año pasado, Luigi Mastroianni, Jr., de la Facultad de Medicina de la Universidad de Pennsylvania, aportó razones para conseguir que se hiciera más investigación sobre la fertilización de los óvulos humanos, basándose en que "el riesgo de devolver un embrión (humano) fertilizado in vitro al útero debe ser examinado con gran cuidado".

Advirtió que, mientras las pruebas con animales son un prerrequisito, no "garantizan completamente que los procedimientos funcionaran igualmente bien en los humanos". En particular, sostenía que la experimentación animal no es hasta ahora estadísticamente válida sobre la cuestión de la normalidad de la descendencia; está hasta aquí limitada a sólo dos especies de laboratorio, y de ellas, ninguna es de primate. Pidió "estudios extensivos adicionales en animales de laboratorio" y con óvulos humanos antes de que la terapia se iniciara sobre seres humanos en los Estados Unidos.

El problema de la seguridad para el feto no está del todo resuelto y sólo puede ser valorado más a fondo a través de pruebas y experimentos adicionales. Las otras cuestiones planteadas por el secretario Califano no gozan de más afortunada respuesta y requieren análisis de muy diferentes tipos. Son lo que Sissela Bok, miembro del comité de revisión, ha llamado cuestiones de "pendiente resbaladiza", es decir, cuestiones acerca de procedimientos desarrollados que pueden tender más o menos inevitablemente a seguir la capacidad recientemente ganada de llevar a cabo la fertilización humana externa. Hay varias posibilidades: (1) fertilización en la que intervienen ovocitos y espermatozoides no obtenidos de parejas casadas; (2) manipulación, además de simple implantación, del embrión antes de ser implantado; (3) trasplante de embriones al útero de una mujer distinta de la que proporcionó el ovocito; (4) ectogénesis, o mantenimiento de los embriones in vi-

FASES SUCESIVAS de la división de los óvulos humanos fertilizados en el laboratorio, según se observa en cuatro microfotografías realizadas por Edwards. Al aumentar el número de células, llamadas blastómeros, éstos se hacen más pequeños, de modo que el tamaño total del agregado celular no aumenta al principio. En el estadio de cuatro células (*arriba a la izquierda*) aún se ven los espermatozoides dentro de la zona pelúcida. La fase siguiente presenta ocho células (*arriba a la derecha*). En la fase de mórula, entre 16 y 32 células, las fronteras entre blastómeros son menos precisas (*abajo a la izquierda*). En el estadio de blastocisto (*abajo a la derecha*) ha desaparecido la zona pelúcida. Las células se han dividido para formar una masa celular interna y un trofoblasto, dejando una cavidad.

OCURRE LA IMPLANTACION cuando el blastocisto se fija al endometrio, erosiona su epitelio y llega a quedar englobado en la pared uterina. La microfotografía, perteneciente a la colección del Departamento de Embriología de la Institución Carnegie de Washington, muestra un blastocisto parcialmente implantado unos siete días después de la fertilización normal (no externa). Tal como se indica en el dibujo adjunto, el embrioblasto se ha convertido en un disco embrionario de dos capas: el epiblasto, que formará las capas embrionarias denominadas ectodermo y mesodermo y el endodermo.

tro más allá de la fase de implantación, y (5) trasplante de partes embrionarias a otros individuos. Consideremos estas cuestiones con especial referencia a su inmediatez, su razón de ser y sus consecuencias verosímiles.

Con respecto a la fertilización no conyugal, cualquier fuente sana de esperma humano tiene todas las posibilidades para ser tan efectiva como cualquier otra a la hora de fertilizar el óvulo y dar lugar a un embrión listo para implantar. La consecuencia biológica sería la misma que la de la inseminación artificial con espermatozoides de otro que no fuera el marido, lo que ya se practica ampliamente como remedio para la infertilidad masculina. Los aspectos éticos y legales deberían ser también muy similares, con tal que el óvulo fertilizado se implantara en la mujer que lo suministró. Realmente el procedimiento ofrecería una oportunidad para una reproducción más selectiva si el esperma donante fuera suministrado por fuentes "deseadas", pero la oportunidad no es mayor que en el caso de la inseminación artificial y las dificultades técnicas son superiores. (Existe también la posibilidad de una reproducción menos selectiva, que sería el caso de fuentes de esperma deliberadamente escogidas al azar para asegurar el anonimato.) Las beneficiarias inmediatas de la fertilización externa no conyugal serían mujeres con los oviductos bloqueados que no tengan un marido fértil; este procedimiento posibilitaría a tal mujer tener un niño que, biológicamente, por lo menos la mitad sería suya.

Con respecto a la segunda cuestión, la fertilización externa sí que proporciona un primer paso técnico hacia la manipulación del desarrollo humano precoz. La observación de los primeros estadios del desarrollo humano ha sido muy limitada por la gran escasez del material apropiado. La experimentación controlada ha estado confinada en gran parte a la fertilización externa y al breve mantenimiento necesario para llevar el cigoto a la fase adecuada para su implantación. Con este éxito, sin embargo, la cuestión del secretario Califano acerca de la posibilidad de "intentos de controlar la estructura genética de la descendencia" deviene por lo menos pertinente.

La intervención genética es una de las posibilidades experimentales que viene a la mente inmediatamente debido a la reciente discusión acerca de las técnicas sobre el ADN recombinante para cambiar el contenido del material genético y la posibilidad teórica de "clonar" seres humanos introduciendo toda la dotación cromosómica de una persona en un

óvulo enucleado. Si se desarrollan medios adecuados para la transferencia de ADN extraño a células humanas, o si las técnicas se desarrollan para introducir núcleos en óvulos enucleados de mamíferos, el procedimiento de Edwards y Steptoe proporcionaría la puerta de entrada para aplicar dichas técnicas a los estadios precoces del desarrollo humano. Los pasos previos todavía por dar no son, sin embargo, simples técnicamente, y los óvulos humanos, dados todos los problemas éticos y logísticos vinculados a su empleo, si apenas son material favorable sobre el que emprender los necesarios estudios.

No obstante, recientemente han aparecido los primeros informes de transferencias con éxito de ADN recombinante a cultivos de células de mamíferos y de renucleación de un óvulo humano con el núcleo de 46 cromosomas de una célula espermática precursora. Ambas maniobras son concebiblemente aplicables a la intervención en la constitución genética humana. La ventana abierta por la fertilización externa en la hasta ahora caja negra que rodeaba el desarrollo humano precoz es, por consiguiente, un paso relevante, aunque no grande, hacia el control de "la estructura genética de la descendencia". Pero el nuevo potencial en esta dirección es con probabilidad más útilmente estudiado en especies no humanas, sobre todo en aplicaciones a la reproducción animal.

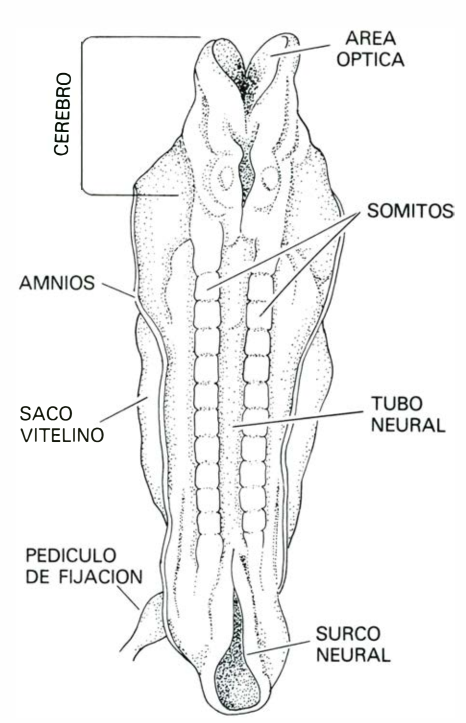
La tercera cuestión tiene que ver con la maternidad subrogativa: el trasplante de un óvulo fertilizado derivado de una mujer al útero de otra. Esta posibilidad ha sido ya comprobada en otras especies de mamíferos y parece que hay pocos obstáculos teóricos o técnicos para llevarlo a cabo en seres humanos. El problema técnico es simple cuestión de armonizar el estadio de desarrollo del embrión extraño con el estadio reproductor del receptor. El estadio de desarrollo ha sido controlado en embriones animales congelándolos, cosa que no parece tener serios efectos nocivos. La adecuada receptividad del útero en una mujer puede predecirse siguiendo su ciclo menstrual y puede controlarse en cierta medida con la administración de hormonas.

Si se desarrolla un consenso ético y social favorable, el papel de nodriza (aceptado en muchas culturas) podría probablemente extenderse bastante rápidamente para incluir el parto subrogativo. Los aspectos sociales, emocionales y éticos de esta práctica reclaman claramente una cuidadosa valoración antes de que la investigación y los ensayos avancen en esta dirección. Debería tomarse en consideración no sólo el valor práctico del procedimiento, sino también el hecho de que probablemente hay mucho que aprender acerca de la génesis de importantes características humanas a través del estudio de la descendencia sometida a ambientes maternos no relacionados genéticamente.

En cuanto a la cuarta cuestión, sobre la continuación del desarrollo externo de embriones humanos más allá de la fase de preimplantación, la investigación sobre ectogénesis está teniendo ya lugar en otras especies de mamíferos y se ha registrado un éxito limitado. La razón de esta experimentación animal es conocer mejor el desarrollo precoz de los mamíferos, tanto por sí mismo como por su relación con el desarrollo humano. Puede aprenderse mucho acerca del papel de los factores genéticos y acerca de los mismos procesos del desarrollo a partir de estudios animales, monitorizando embriones enteros de mamíferos y sus partes mientras se desarrollan en un aislamiento externo. Hay algunas cosas, sin embargo, como la sensibilidad humana específica a los fármacos y las funciones cerebrales superiores, que pueden comprenderse sólo parcialmente a través de los estudios sobre animales, y puede obtenerse importante información estudiando embriones humanos mantenidos en cultivo externo más allá de la fase de implantación. Hay igualmente importantes reservas, que discutiré a continuación.

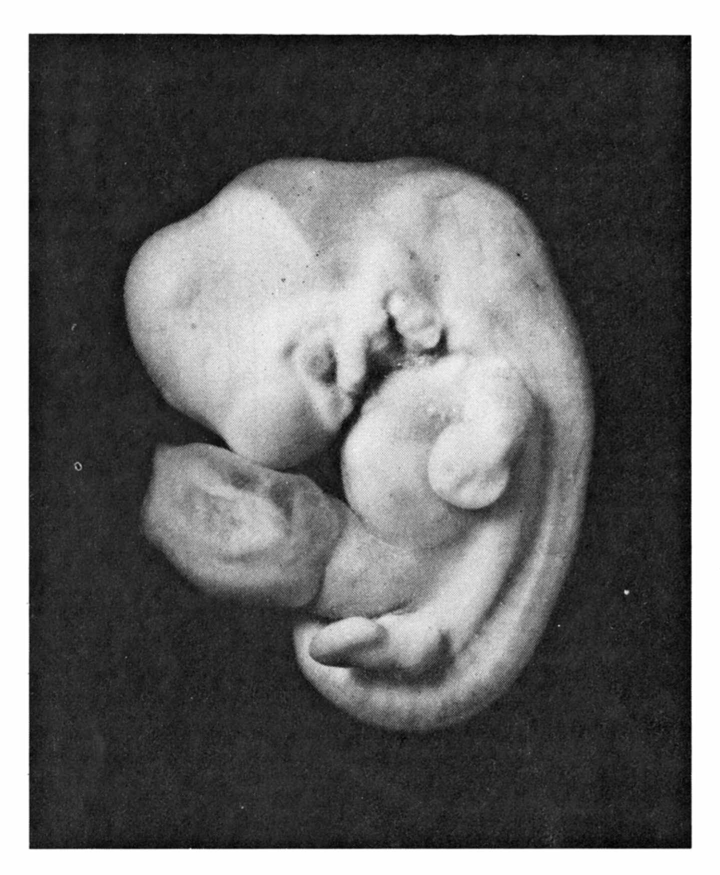
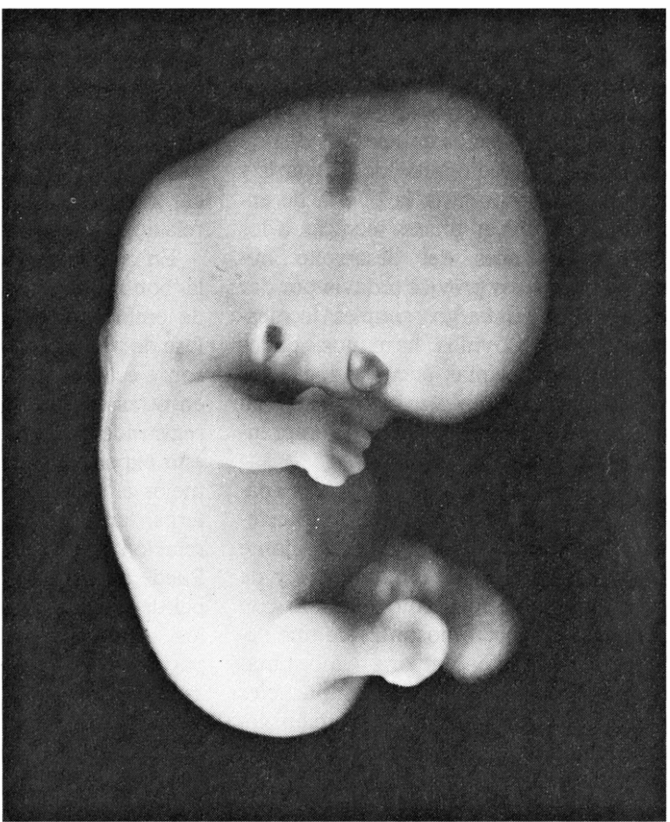
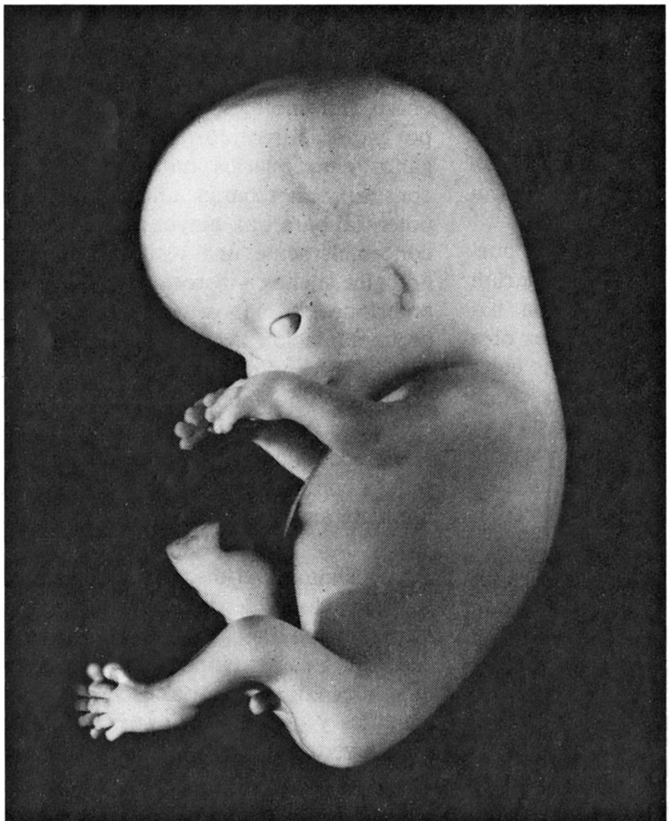
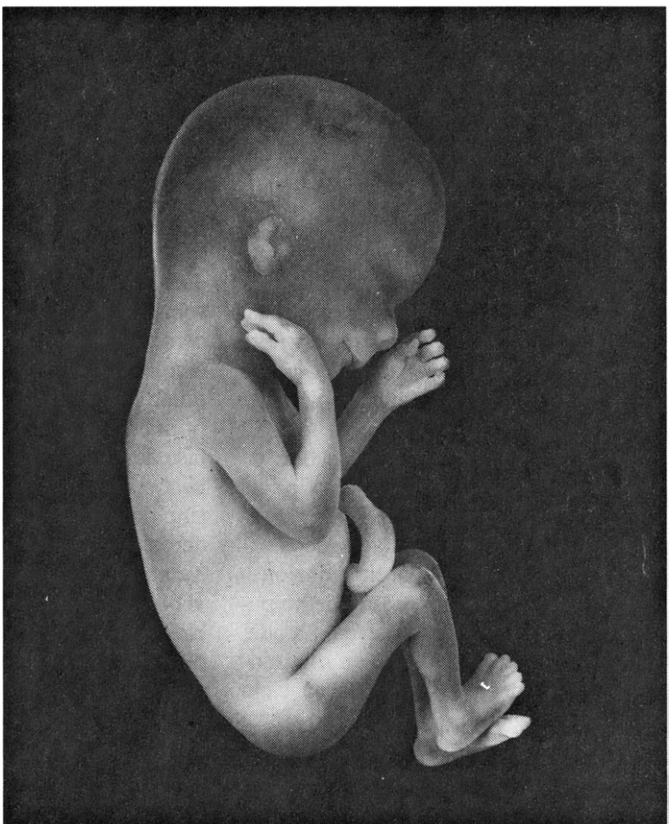
Cualesquiera conclusiones que puedan alcanzarse acerca de la investigación sobre embriones jóvenes, debería hacerse hincapié en que hay muchos obstáculos prácticos importantes ante cualquier intento de prolongar el desarrollo humano precoz externamente hasta una fase de viabilidad independiente y así, de hecho, eliminar el desarrollo intrauterino del ciclo vital humano. Estos verdaderos "niños probeta" y los criaderos humanos imaginados en *Brave New World* de Aldous Huxley no son factibles en el estado actual del conocimiento ni, que se sepa, son éstos los objetivos científicos. Si alguna vez llegan a ser objetivos humanos, lo decidirá el futuro.

La quinta cuestión enfatiza la importancia de la cuidadosa consideración de la cuarta. El valor del trasplante de órganos y tejidos en la terapia de enfermedades degenerativas y traumáticas ha sido demostrado por los espectaculares (si no logrados del todo) esfuerzos quirúrgicos en relación con la piel, los riñones y el corazón. El trasplante procedente de donantes cooperativos y de víctimas de muerte accidental o prematura de otro tipo, está limitado por la disponibilidad de recursos y por el rechazo inmunológico de tejido extraño por parte del receptor. Los tejidos y órganos embrionarios humanos podrían constituir un recurso adicional, con el potencial para una mayor abundancia y concebiblemente una resistencia inmunológica menos vigorosa. El cultivo de embriones humanos desde la fase de preimplantación hasta el estadio de formación precoz de órganos podría proporcionar una fuente terapéutica práctica, pues se ha visto que muchos órganos precoces de otros embriones de mamífero continúan desarrollándose externamente aislados del resto del embrión.



EMBRION JOVEN, en el estadio de 10 somitos alcanzado durante la cuarta semana después de la fertilización, formado por el plegamiento de diferentes capas celulares. Los somitos, ordenados dorsalmente en el embrión, son masas de células que se desarrollarán para formar músculo, elementos esqueléticos y dermis, capa profunda de la piel. (A las 20 semanas, un feto es una persona.)

Hasta ahora he discutido estas cinco cuestiones desde un punto de vista estrictamente biológico, pero claramente cada una de ellas también da lugar a inquietudes éticas importantes. Lo que es científica y técnicamente posible, puede no ser social o individualmente aceptable. Hay controversia sobre problemas tales como "el derecho a la vida", el derecho a la muerte, ingeniería genética y consentimiento responsable de la experimentación humana, porque en cada una de estas áreas los conocimientos y capacidades biomédicos en progreso han tropezado con ideas políticas, sociales, éticas y religiosas profundamente arraigadas. El debate crece y declina conforme los nuevos conocimientos agitan proble-

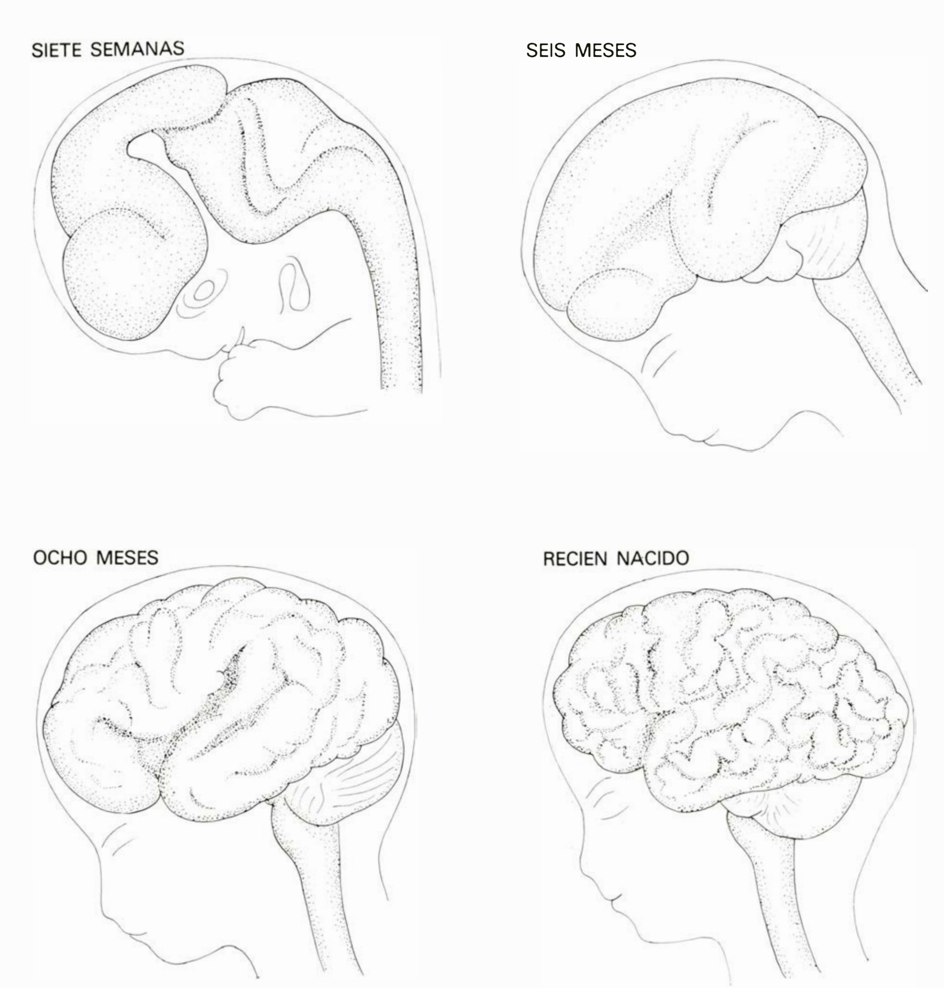
EMBRIONES MADUROS Y FETOS de la colección Carnegie. Los especímenes son embriones de unas seis (*arriba a la izquierda*), siete (*arriba a la derecha*) y ocho (*abajo a la izquierda*) semanas después de la ovulación y un feto (*abajo a la derecha*) de unas 13 semanas; miden (de la coronilla a las nalgas) unos 12, 18, 31 y 100 milímetros de longitud. En el embrión de seis semanas aún son visibles los vestigios de los somitos a lo largo de la espina dorsal. En los esbozos de las extremidades se han formado las primitivas paletas de la mano y del pie; una prominencia por encima del cordón umbilical (que se extiende hacia la izquierda) indica el corazón y el hígado. A las siete semanas, el comienzo de la formación del ojo es evidente y la cabeza está más claramente definida. Los elementos de los brazos y las piernas están también definidos y las estrías representan los dedos en desarrollo. A las ocho semanas, los hemisferios cerebrales han crecido, los pliegues palpebrales empiezan a cubrir los ojos y la nariz se hace prominente. Las partes de las extremidades están bien definidas y diferenciados los dedos de manos y pies. Las diferencias en los genitales externos permite la identificación sexual. El feto de 13 semanas es claramente reconocible como humano por un lego. La cabeza es aún muy grande. La oreja, nariz y boca se aproximan en apariencia a los de un niño a término. Los ojos han sido cubiertos por los pliegues palpebrales. (Dep. de Embriología de Carnegie.)

mas particulares. La fertilización humana externa es un avance que con seguridad estimulará el debate general y enfocará la atención en aspectos particulares del mismo. El aspecto que cobra mayor importancia es la cuestión del estadio del desarrollo humano en el cual una "persona", en el sentido ético y legal, empieza a ser tal. Es importante, dados los muy diversos puntos de vista sobre la materia que actualmente se sostienen, comprender lo que puede decirse científicamente acerca de este problema central.

Algunos razonan, sobre bases no científicas, que la génesis de una persona no es de desarrollo, en el sentido de una formación gradual durante un período de tiempo; que el momento de la fertilización –concepción– marca el origen de una persona porque en ese momento las contribuciones materna y paterna se combinan estableciéndose una nueva individualidad. Científicamente, no obstante, la fertilización en sí misma es un proceso en el tiempo. Acontecimientos sucesivos, como los que siguen, se distinguen en el proceso: contacto del espermatozoide con la superficie del ovocito; "activación" de la corteza o capa externa del ovocito; formación del cono de fertilización, extrusión del citoplasma hacia el espermatozoide; entrada del espermatozoide en el citoplasma del ovocito; terminación de la meiosis y maduración final del núcleo del óvulo; formación de los pronúcleos del espermatozoide y del óvulo; fusión de los pronúcleos. Estos acontecimientos, que ocurren en unas horas, culminan con la primera división, la división del cigoto para formar dos células.

Una fase particular, como es la entrada de la cabeza del espermatozoide o la fusión de los pronúcleos, podría pensarse que constituye el momento del establecimiento de la individualidad, pero ninguno de ellos es un paso esencial para el desarrollo subsiguiente; en varios ovocitos animales son suficientes otros tipos de activación de su corteza para iniciar el desarrollo sin una contribución paterna (partenogénesis). E incluso cuando los cromosomas paternos están presentes, se ha demostrado que no tienen efecto hasta bien superada la fertilización. Por tanto, la fertilización, como los pasos para la formación del óvulo y del espermatozoide que la preceden y las fases de división que la siguen, es un proceso complejo y continuo en el tiempo. La elección de un momento particular como iniciación de la individualidad es necesariamente arbitrario.

Científicamente hay dos acontecimientos esenciales en la fertilización. En



DESARROLLO DEL CEREBRO correspondiente a cuatro fases embrionarias y fetales. El joven cerebro de forma tubular empieza a alcanzar su forma madura como resultado de una serie de curvas, o flexiones. Las diferentes regiones del cerebro crecen de una manera diferenciada y los hemisferios cerebrales cubren el cerebelo y otras regiones. Aunque la silueta principal del cerebro queda establecida al sexto mes, las circonvoluciones y fisuras de la superficie, que aumentan enormemente el área y volumen de la corteza, no proliferan hasta aproximadamente el octavo mes de vida intrauterina.

primer lugar, la larga interrupción del desarrollo que es característica del óvulo maduro se rompe por una serie de cambios progresivos que inician un nuevo ciclo vital. En segundo lugar, en los seres humanos se forma una nueva constitución genética contenida en dos conjuntos homólogos de cromosomas. La cuestión real es si una "persona" existe tan pronto como empiezan los cambios progresivos del mecanismo de la fertilización. La respuesta científica es negativa, a no ser que se mantenga el ingenuo preformismo que precedió a la moderna embriología: el concepto según el cual un homúnculo –un ser humano completo pero en miniatura– está presente o bien en el óvulo o bien en el espermatozoide antes de la fertilización.

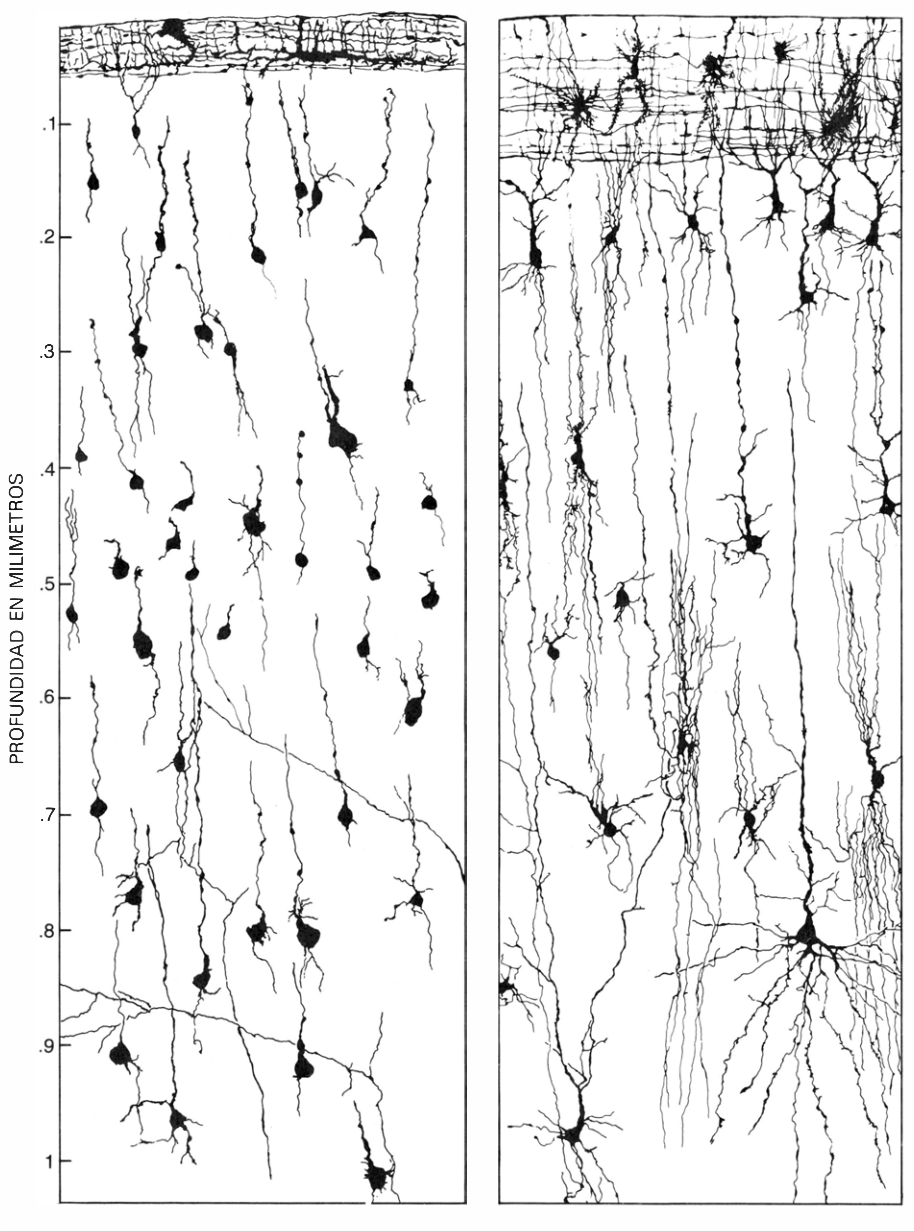
Hoy en día hay pruebas aplastantes de que esta noción es falsa. Lo que está presente en el óvulo fertilizado es información genética, inscrita en lenguaje químico en moléculas de ADN (y en elementos citoplasmáticos formados bajo la influencia del ADN materno). A través de complejos procesos esa información será traducida y elaborada para dar una nueva persona. La evidencia también indica que la traducción y elaboración serán influenciadas por muchos factores en el transcurso del desarrollo y que algunos de estos factores fueron generados en el ovocito mucho antes de la fertilización. Bajo ciertas circunstancias es posible que sean producidos dos individuos por un óvulo fertilizado (formación de gemelos) o incluso que no se produzca ninguna persona verdadera (en el caso de anencefalia, o ausencia de cerebro).

Además, la minuciosa observación de un óvulo justo antes o durante un cierto tiempo después de la fertilización revela pocas de las características por las que reconocemos a una persona. Lo que revela son todas las características de una célula humana individual.

La fertilización no cambia el carácter de célula individual del óvulo. Sólo después de iniciada la división celular, y de

los múltiples y complejos cambios que siguen, el cigoto empieza a desarrollar el conjunto de características de un organismo pluricelular y finalmente de un individuo altamente organizado. Lo que es especial en un óvulo humano es su capacidad, bajo circunstancias favorables, de dar lugar a una persona; no hay pruebas científicas de que se trate realmente de una persona.

Si una persona no puede identificarse en el óvulo, ¿cuándo puede identificarse en el sistema celular en desarrollo? Es importante reconocer que hay dos aspectos del concepto de persona en cuanto se relaciona con el desarrollo humano: el de la percepción externa y el de la experiencia interna. Los dos aspectos no se corresponden necesariamente en el curso del desarrollo; sin embargo, ambos tienen una importante influencia en la cuestión del momento de aparición de la cualidad interna de persona. El aspecto externo es diagnosticable objetivamente; el interno es subjetivo y puede, en el mejor de los casos, tan sólo inferirse objetivamente. El juicio externo puede depender de criterios estructurales y de comportamientos asociados con la cualidad de persona. Cualquier juicio sobre la cualidad interna de persona resulta frustrado, porque el sentir, el dolor y la capacidad para el conocimiento del yo son esencialmente imposibles de establecer sin algunos medios de comunicación.



NEURONAS DE LA CORTEZA VISUAL de fetos abortados de 25 semanas (*izquierda*) y 33 semanas (*derecha*) se representan en estos esquemas realizados a partir de microfotografías registradas por Dominick P. Purpura, del Albert Einstein College of Medicine. Se tiñó el tejido con el método de Golgi. La profundidad respecto a la superficie está representada en milímetros. Sólo se recoge la mitad superior de la corteza en el especimen de más edad (*derecha*); en realidad la corteza visual prácticamente dobla el espesor de 25 a 33 semanas. La diferencia obvia entre las dos muestras es el gran aumento en número y longitud de las dendritas, los procesos más cortos que median la transmisión de los impulsos nerviosos entre las neuronas. Las dendritas son también más maduras, mostrando (a grandes aumentos) numerosas "espinas", donde tiene lugar la transmisión de los impulsos nerviosos.

Dependiendo totalmente de signos externos, por tanto, se pueden distinguir diversas fases escalonadas del desarrollo humano: la fase celular o preembrionaria, la fase embrionaria, la fase fetal y finalmente la fase de autonomía vital, en particular con respecto a la respiración. Estas son fases escalonadas de forma continua, pero conviene distinguirlas entre sí. Las dos primeras fases, preembrionaria y embrionaria, pueden reconocerse como ser humano sólo por expertos; y cuando los expertos reconocen el carácter humano de estos estadios, se basan en criterios diferentes de los que generalmente se aplican a una persona. La razón estriba en que las propiedades de los seres humanos (así como las propiedades de la vida) no son lo mismo que las propiedades de las personas. Los estadios preembrionario y embrionario son vivos y son humanos, pero no son externamente reconocibles como persona. (Lo mismo es verdad, claro está, para las células y tejidos que son separados de un adulto: pueden permanecer vivos y humanos, pero no constituyen una persona.)

El óvulo fertilizado y el agregado celular que origina antes de la implantación tienen la capacidad, bajo circunstancias favorables, de dar lugar a una persona, pero a menudo no lo consiguen, ni siquiera bajo condiciones naturales. Sólo después de la implantación, el aumento de tamaño, la génesis de la forma y la aparición de estructuras y órganos rudimentarios marcan la transformación del agregado celular en un embrión. Hasta después de unas ocho semanas de desarrollo, el embrión no dispone de características que empiecen a ser externamente reconocibles como humanas por los inexpertos. A las 20 semanas, el aspecto externo de lo que ahora es un feto es tal que hasta los no expertos lo considerarían probablemente como una persona en desarrollo.

Por tanto, la transformación de un embrión en feto, aunque no tiene unos límites muy precisos, es una transformación de su identificación externa sólo como humano en su identificación cada vez mayor como persona. No se sabe si estos cambios percibidos externamente van acompañados de otros internos que establecen algunos aspectos interiores de la cualidad de persona, pero el estado del desarrollo del cerebro hace que esto sea poco probable hasta por lo menos una fase tardía.

En los últimos años se ha registrado un progreso considerable en la comprensión del desarrollo y actividad del cerebro. Varias cosas parecen razonablemente claras. No sólo el comportamiento, sino también el estado de conciencia complejo, están íntimamente ligados a la función cerebral. Un aspecto esencial de la función cerebral humana es el número de neuronas (miles de millones) y las ricas conexiones entre neuronas (muchísimas más) de la corteza cerebral. El número de neuronas corticales continúa aumentando dentro de los seis meses de vida fetal, pero las conexiones entre estas neuronas aumentan enormemente, no sólo bien pasados los seis meses fetales, sino también durante cierto número de años de vida postnatal. El estado inmaduro del desarrollo del cerebro durante la vida fetal se corresponde bien con las observaciones sobre el comportamiento fetal. Sensibilidad, diversos reflejos e incluso movimiento espontáneo son manifiestos en el último tercio del embarazo, pero hay una mínima evidencia de función cognoscitiva incluso en el nacimiento.

No obstante, los cambios externos durante el final de la vida fetal son significantes para definir la génesis de una persona, ya que evocan reconocimiento y emociones afectivas en otras personas –y estas emociones importan, porque la cualidad de persona es un status social así como un estado del individuo. En estos términos, el período fetal es aquel durante el cual aparecen los primeros signos externos de la cualidad de persona, según se aprecia por el aumento de empatía emocional provocada en otras personas. Cuando esta empatía se hace suficientemente grande, el feto se ha convertido en un niño –una persona– independientemente de si está aún en el útero o de si ha sido dado ya a luz. Esta determinación es independiente tanto del momento de nacer, que puede variar por razones biológicas y médicotecnológicas, como del aspecto interno de la cualidad de persona, cuya determinación está llena de dificultades. El propósito debería ser el de definir la cualidad de persona de tal forma que se asegure que el status se confiere bastante antes de que pueda haber cualquier posibilidad razonable de que el aspecto interno ha sido de hecho alcanzado.

Al valorar la fertilización externa, por tanto, el punto más importante es que, por lo menos hasta las ocho semanas, puede considerarse con seguridad que el embrión humano carece aún de los dos aspectos esenciales de la cualidad de persona: identificación afectiva por otras personas y conocimiento interno consciente. El procedimiento Edwards-Steptoe no manipula personas sino células humanas. Las fases involucradas no sólo son prepersonas sino que también son preembriones. Son preembriones porque las células separadas de los estadios iniciales de división de los huevos de mamíferos no humanos (y muy probablemente de huevos humanos) pueden cada una dar lugar a un embrión completo en los casos gemelares, y porque pueden extirparse células de los grupos celulares iniciales (o pueden combinarse dos grupos) e incluso dar lugar a un solo embrión. Por tanto, en los estadios celulares anteriores a la implantación, aún no ha sido establecido un embrión individual. Puesto que ni siquiera en la fase embrionaria no han aparecido aún los criterios que definen la cualidad de persona, el procedimiento de Edwards-Steptoe permite un amplio margen de seguridad para asegurar que las personas no están siendo manipuladas y que los derechos de las personas no se están violando.

Esto no excluye la preocupación por las consecuencias en caso de que los productos del procedimiento se implanten o se nutran de otro modo, de manera que puedan dar lugar a una persona. Cualquier célula o grupo de células que sea progenitor de una persona requiere protección frente a intervenciones o influencias que puedan limitar la calidad de la vida de esta persona. Por tanto, hay buenos fundamentos para insistir que los ovocitos a implantar, para continuar el desarrollo, deben ser valorados cuidadosamente frente a cualquier posible deterioro. Esto no quiere decir, sin embargo, que los huevos que no se implanten tengan un status especial más allá del status especial normalmente asignado a células, tejidos o partes humanas. La capacidad de producir una persona bajo circunstancias apropiadas no debería confundirse con la realización de esta capacidad.

Una cuestión importante, que ciertamente surgirá antes de mucho, es si debería emprenderse la investigación para prolongar el desarrollo externo de los estadios celulares del desarrollo humano hasta fases embrionarias. Hay preocupación de que tales estudios, si fueran suficientemente extensos, podrían conducir a la ectogénesis completa, pero esto actualmente debe considerarse fuera de alcance, incluso si se deseara. Lo que no está fuera de alcance es una mayor comprensión de las condiciones necesarias para transformar la fase celular en fase embrionaria, que ocurre mucho antes de la consecución de los criterios que definen la cualidad de persona; excepto la ectogénesis, estas fases no pueden probablemente dar lugar a una persona. Una vez estas fases son alcanzadas externamente, no obstante, podrían proporcionar una fuente para el desarrollo continuado de rudimentos de órganos, mediante técnicas que ya han sido elaboradas para mamíferos de laboratorio. Pero estos rudimentos, como los embriones enteros de los que proceden, no podrían de ninguna manera dar lugar a personas. La política con respecto a estas posibilidades precisa una profunda consideración.

Debemos agradecer la aparición de nuevos planteamientos para discutir estos asuntos en la Ethics Advisory Board y en la Office of Technology Assessment, que tiene entre manos preparar un juicio de genética aplicada. Es de esperar que esta consideración, con una completa discusión pública, produzca directrices que puedan ser revisadas y modificadas a medida que los problemas surjan y cambien. Dada la complejidad de los problemas y las profundas reacciones sociales y emocionales que engendran, parece lógico considerar primero el establecimiento de pautas para los problemas que están al margen, tanto por su inmediatez como por nivel de consenso en que se hallan. Un ataque frontal al complejo entero sólo generaría una frustrante controversia entre opuestos puntos de vista generales.

Como ejemplos, los estudios de laboratorio sobre fertilización humana y estadios de preimplantación requieren una consideración inmediata y deberían trazarse directrices apropiadas para la obtención de óvulos y la utilización de los productos de fertilización. Capítulos menos apremiantes en el orden del día son los referentes al status de las células embrionarias, tejidos y órganos, se obtengan de abortos o de fases de preimplantación desarrolladas posteriormente en el laboratorio. Otros aspectos que merecen atenderse son las diversas manipulaciones de las fases de preimplantación antes de la implantación, entre las que se cuentan la fusión celular, la intervención genética y la obtención de clones.

El colocar estos puntos en el orden del día no presupone su conveniencia; pero su consideración puede ayudar a establecer el límite apropiado más allá del cual la política actual no debería ser tolerante. Este límite quizá se coloque lejos de la producción externa de fases sensibles del desarrollo o de la producción de personas que están limitadas o alteradas en cuanto a la calidad de vida.

El factor de crecimiento nervioso

Esta proteína desempeña un papel esencial en la formación del sistema nervioso. Y se ha utilizado como un "hilo de Ariadna" para explorar los mecanismos de crecimiento y diferenciación de la propia neurona

Rita Levi-Montalcini y Pietro Calissano

El sistema nervioso humano es una vasta red de varios miles de millones de neuronas –células nerviosas– dotadas de la notable capacidad de recibir, almacenar y transmitir información. A fin de comunicarse entre sí y con células no nerviosas, las neuronas cuentan con largas extensiones denominadas axones, de alguna manera comparables a cables de conducción eléctrica. Pero, a diferencia de los cables, los axones son estructuras cilíndricas llenas de líquido que no sólo transmiten señales eléctricas, sino que transportan también al cuerpo celular, o desde él, sustancias nutritivas y otras moléculas esenciales. Aún quedan muchas cuestiones básicas por resolver en relación con los mecanismos que rigen la formación de esta intrincada red celular. ¿Cómo tiene lugar la diferenciación de las células en miles de tipos distintos? ¿Cómo establecen los axones sus conexiones específicas (sinapsis) con otras neuronas y con células de otros tejidos? ¿Cuál es la naturaleza de los mensajes químicos enviados y recibidos por las neuronas una vez se han establecido las conexiones sinápticas?

Este artículo describe algunas de las principales características y efectos de una proteína llamada factor de crecimiento nervioso (FCN), la cual ha permitido inducir y analizar en condiciones muy favorables algunas etapas cruciales de la diferenciación de las neuronas, tales como el crecimiento y maduración de los axones y la síntesis y liberación de neurotransmisores, los portadores de los mensajes químicos. El descubrimiento del FCN ha suscitado una intensa investigación de otros factores de crecimiento específicos, lo que ha conducido al aislamiento y caracterización de una serie de proteínas capaces de incrementar el crecimiento de diferentes líneas celulares.

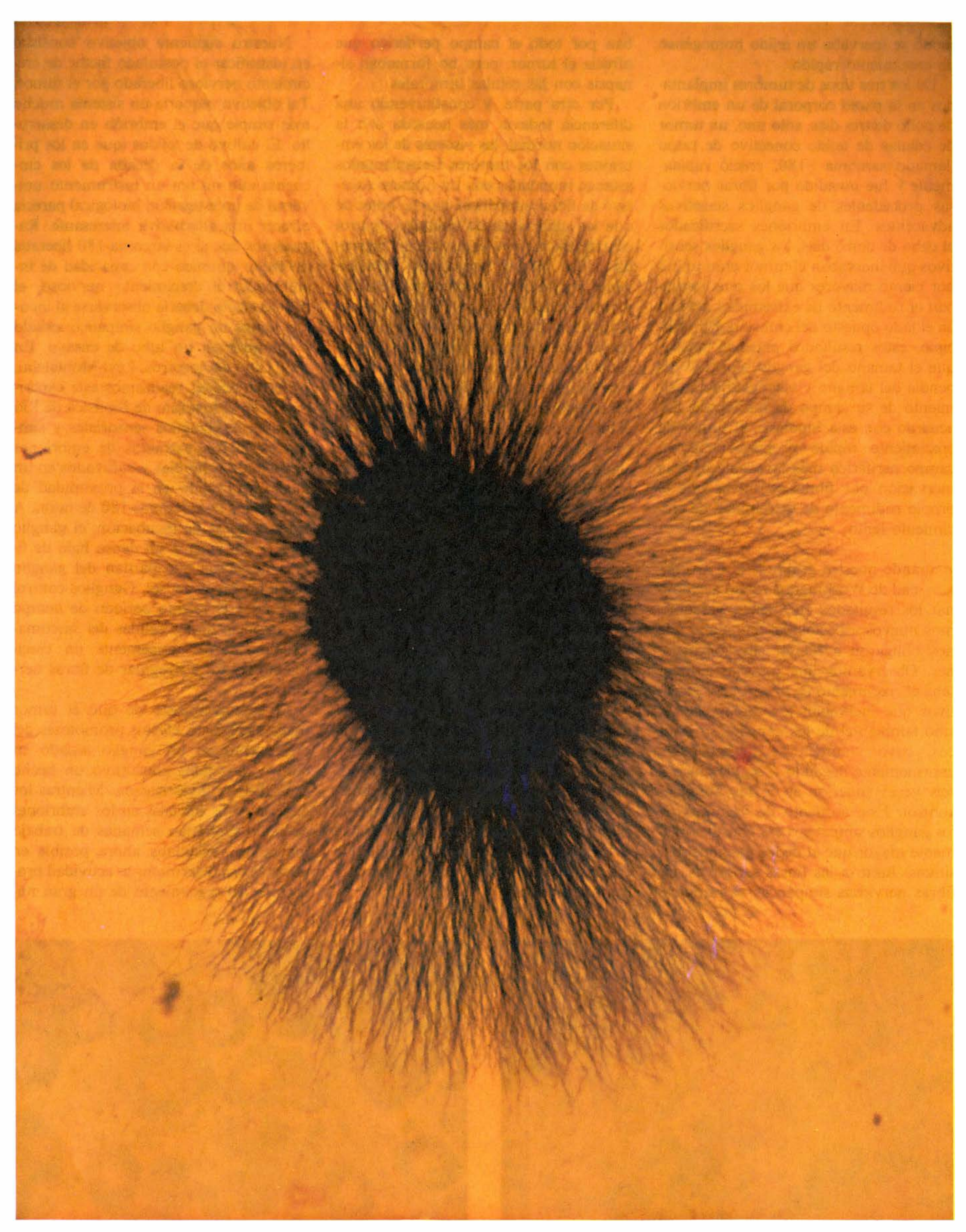
El sistema nervioso periférico de los vertebrados comprende tres clases de células nerviosas: las neuronas sensitivas que transmiten impulsos desde los receptores sensitivos al cerebro; las neuronas motoras que inervan los músculos estriados, o esqueléticos; y las neuronas del sistema nervioso vegetativo, que regulan la actividad funcional del sistema circulatorio, los órganos, las glándulas y la musculatura lisa (como la del intestino). Las neuronas del sistema nervioso vegetativo son de dos clases: simpáticas y parasimpáticas. Las neuronas sensitivas y algunas neuronas simpáticas están situadas en cadenas de ganglios que flanquean longitudinalmente la médula espinal. Por ser estas neuronas particularmente accesibles a las manipulaciones experimentales, la mayor parte de la investigación del desarrollo del sistema nervioso a nivel celular se ha centrado en el mecanismo mediante el cual las fibras nerviosas que se proyectan desde los ganglios sensoriales y simpáticos establecen conexiones con sus correspondientes órganos efectores.

En la primera mitad de este siglo, la nueva ciencia de la embriología experimental parecía ofrecer la mejor aproximación al estudio de las uniones íntimas que conectan las neuronas periféricas en crecimiento y sus órganos efectores. Ross G. Harrison, de la Universidad de Yale, puso a prueba el sistema nervioso de larvas de anfibio, proponiéndole la resolución de un problema con el que normalmente nunca se enfrentan tales sistemas; se trataba de inervar las extremidades o los órganos transplantados de otras especies. Estaba interesado en observar cómo los ganglios sensitivos y simpáticos, al enviar fibras nerviosas a los "campos de inervación" periféricos, se ajustarían a las diferentes dimensiones y configuraciones de los órganos extraños.

Los resultados de Harrison demostraron que el sistema nervioso de los anfibios en desarrollo es extraordinariamente flexible en su adaptación a situaciones nuevas, hasta el punto de acelerar el crecimiento de las fibras nerviosas del animal receptor para armonizar con la mayor velocidad de crecimiento de la extremidad de la especie dadora. Llegó a la conclusión de que el sistema nervioso embrionario es muy sensible a influencias ejercidas por el campo periférico. Tales influencias no las produce exclusivamente el campo periférico de la propia especie, puesto que aparecen también al transplantar órganos o extremidades rudimentarias procedentes de otras especies.

Viktor Hamburger, de la Universidad de Washington, amplió los estudios de Harrison eligiendo para sus experimentos el embrión de pollo, que, aunque posee un sistema nervioso más complejo que el de los anfibios, ofrecía algunas ventajas para el análisis experimental. Los centros nerviosos están mejor delimitados y su fuerte afinidad para la tinción argéntica permite al investigador observar más fácilmente las estructuras nerviosas. Hamburger injertó extremidades rudimentarias en embriones de pollo muy al principio de la embriogénesis y observó cómo el campo periférico modificado era inervado por fibras sensitivas y simpáticas. Desafortunadamente, las respuestas fueron a menudo tan complejas que hacían muy difícil la interpretación.

En 1948, en un esfuerzo para obtener resultados más directos, Elmer D. Bueker, de la Universidad Georgetown, modificó el enfoque experimental de Hamburger. Tuvo la idea ingeniosa de reemplazar la extremidad rudimentaria de un embrión de pollo por un fragmento de un tumor de ave o mamífero. Las células tumorales estaban todas indiferenciadas y proporcionaban un campo periférico homogéneo en contraposición a las células del rudimento de extremidad normal, destinadas a diferenciarse en múltiples tejidos. Los experimentos de Bueker deberían, pues, revelar de qué



HALO DE FIBRAS NERVIOSAS irradiado por un ganglio simpático, aislado a partir de un embrión de pollo e incubado durante 12 horas en un medio semisólido en presencia del factor de crecimiento nervioso (FCN). El factor fue obtenido originalmente de un tumor de ratón denominado sarcoma −180, pero ha sido aislado ulteriormente partiendo de una gran variedad de células normales y neoplásicas (cancerosas) que lo segregan en pequeñísimas cantidades. En 1953, uno de los autores (Levi-Montalcini) y Hertha Meyer observaron que el FCN podía inducir el crecimiento de fibras nerviosas en ganglios simpáticos aislados mantenidos en cultivo de tejidos. Tal descubrimiento condujo finalmente al aislamiento del FCN y a la determinación de su estructura química. La micrografía representa un ganglio con una ampliación de 135 diámetros y fue obtenida por J. S. Chen.

modo se inervaba un tejido homogéneo de crecimiento rápido.

De los tres tipos de tumores implantados en la pared corporal de un embrión de pollo de tres días, sólo uno, un tumor de células de tejido conectivo de ratón llamado sarcoma −180, creció rápidamente y fue invadido por fibras nerviosas procedentes de ganglios sensitivos adyacentes. En embriones sacrificados al cabo de cinco días, los ganglios sensitivos que inervaban el tumor eran un 33 por ciento mayores que los que inervaban el rudimento de extremidad normal en el lado opuesto del embrión. En principio, estos resultados parecían sugerir que el tamaño del ganglio sensitivo dependía del tamaño y velocidad de crecimiento de su campo de inervación. De acuerdo con esta hipótesis, el tumor de crecimiento rápido proporcionaba un campo periférico más favorable para la inervación por fibras sensitivas que el propio rudimento de extremidad de crecimiento lento.

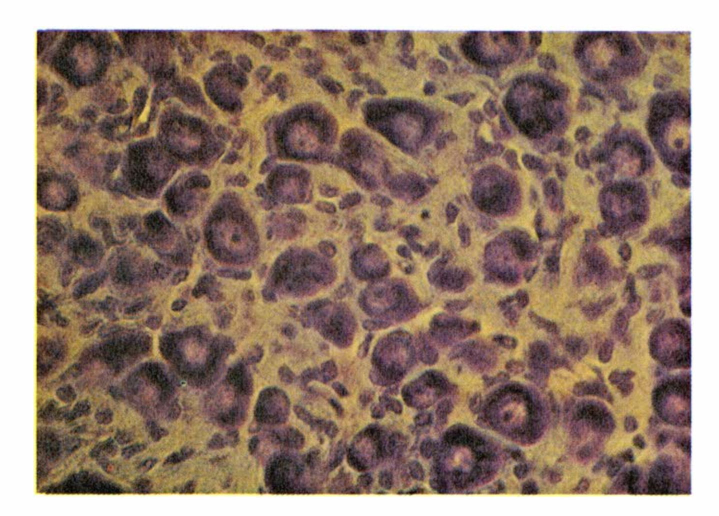
Cuando nuestro grupo de la Universidad de Washington volvió a examinar los resultados de Bueker, descubrimos nuevos aspectos del fenómeno que nos obligaron a revisar sus conclusiones. Observamos que no sólo aumentaba el crecimiento de los ganglios sensitivos que inervaban los sarcomas-180, sino también el de los ganglios simpáticos, cuyo volumen se incrementaba enormemente llegando a ser de cinco a seis veces superior al de los animales control. Este aumento de volumen de los ganglios simpáticos era considerablemente mayor que el de los ganglios sensitivos. Junto a las fibras sensitivas, las fibras nerviosas simpáticas se ramificaban por todo el campo periférico que ofrecía el tumor, pero no formaban sinapsis con las células tumorales.

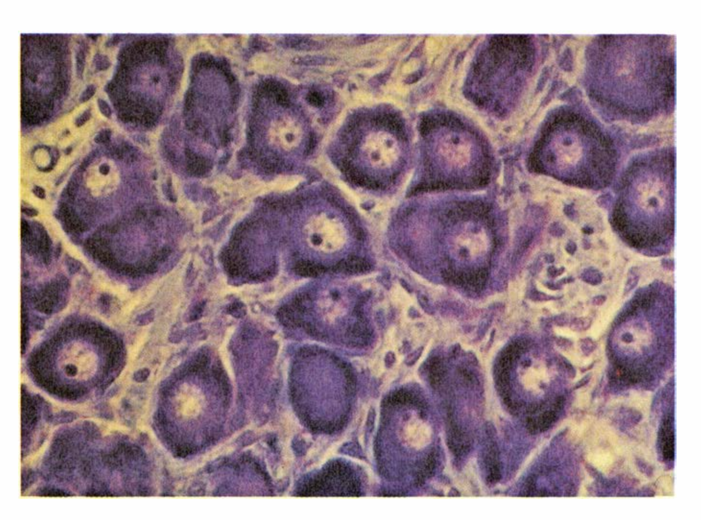
Por otra parte, y constituyendo una diferencia todavía más acusada con la situación normal, las vísceras de los embriones con los tumores transplantados estaban inundadas con un número excesivo de fibras simpáticas mucho antes de que los embriones de animales control estuviesen incluso escasamente inervados. Las fibras simpáticas invadían venas pequeñas y grandes, dificultando y obstruyendo a veces el flujo de sangre. Estos efectos extraordinarios sugirieron que el exceso de crecimiento de los ganglios simpáticos era algo más que una simple respuesta al rápido crecimiento del campo periférico proporcionado por el tumor, como había propuesto Bueker. Más bien parecía que el tumor liberaba algún factor químico que induciría el extraordinario crecimiento de los ganglios simpáticos y la ramificación exuberante de sus fibras nerviosas.

Para probar la validez de esta hipótesis fueron transplantados los sarcomas-180 en las membranas respiratorias del huevo de pollo, que están irrigadas con vasos sanguíneos del embrión. El tumor y el embrión de pollo compartían, por tanto, el mismo flujo de sangre, aunque no estaban en contacto directo. Observamos que, cuando el tumor era transplantado en las membranas respiratorias, inducía los mismos efectos promotores del crecimiento de los ganglios simpáticos que al ser implantado directamente en el embrión, lo que proporcionaba una prueba convincente de que el tumor liberaba un factor soluble que era transportado por la corriente sanguínea al embrión.

Nuestro siguiente objetivo consistió en identificar el postulado factor de crecimiento nervioso liberado por el tumor. Tal objetivo requería un sistema mucho más simple que el embrión en desarrollo. El cultivo de tejidos (que en los primeros años de la década de los cincuenta aún no era un instrumento universal de investigación biológica) parecía ofrecer una alternativa interesante. Razonamos que si el sarcoma-180 liberaba un factor químico con capacidad de incrementar el crecimiento nervioso, el mismo efecto debería observarse al incubar juntos un ganglio simpático aislado y el tumor en un tubo de ensayo. En 1953, uno de nosotros, Levi-Montalcini, y Hertha Meyer realizamos este experimento en el Instituto de Biofísica de Río de Janeiro. Ganglios sensoriales y simpáticos fueron extraídos de embriones de pollo de ocho días y cultivados en un medio semisólido en la proximidad de fragmentos de sarcomas-180 de ratón. A las diez horas de incubación, el ganglio aislado dio lugar a un denso halo de fibras nerviosas que partían del ganglio como los rayos del sol. Ganglios control cultivados el mismo período de tiempo en ausencia de las células del sarcoma-180 mostraron únicamente un crecimiento escaso e irregular de fibras nerviosas.

El descubrimiento de que el tumor podía ejercer sus efectos promotores del crecimiento en un ganglio aislado en cultivo de tejidos constituyó un hecho crucial de la investigación. Mientras los experimentos iniciales en los embriones de pollo requerían semanas de trabajo meticuloso, resultaba ahora posible en pocas horas determinar la actividad promotora del crecimiento de un gran nú-





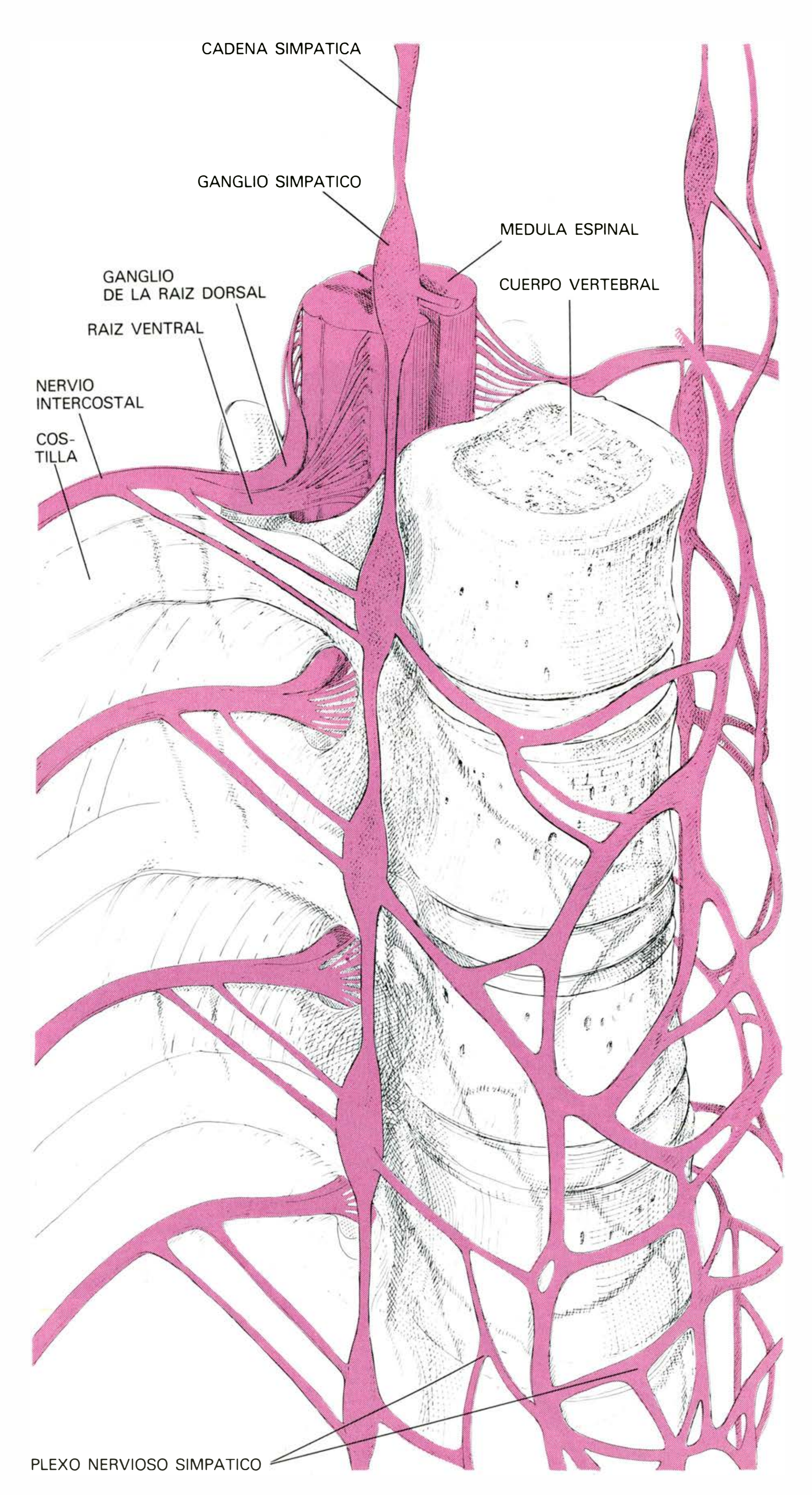
INCREMENTO DEL TAMAÑO Y NUMERO de neuronas en los ganglios simpáticos de un ratón tratado con FCN. El tratamiento da lugar a un aumento del volumen de los ganglios. Las micrografías representan secciones a través de dos ganglios, un ganglio aislado de un ratón de tres semanas de edad inyectado diariamente con solución salina (*izquierda*) y el otro de un ratón de la camada inyectado con FCN (*derecha*). Las neuronas del ratón tratado con FCN tienen mayor volumen. Las células ganglionares de animales tratados con FCN muestran mayor afinidad por el azul de toluidina.

mero de tejidos, líquidos orgánicos y substancias químicas. Resultaba, además, posible intentar aislar el factor de crecimiento nervioso a partir de nuestro sistema simplificado de cultivo.

Un bioquímico, Stanley Cohen, se unió a nuestro grupo de la Universidad de Washington para emprender la tarea de la identificación química del agente activo. Muy pronto consiguió aislar el factor de crecimiento en una fracción de las células tumorales que contenía proteínas y ácidos nucleicos. Aunque este hallazgo abrió el camino para un análisis bioquímico más preciso, no se hubieran realizado progresos notables a no ser por un hecho fortuito acaecido dos años más tarde. Para determinar si el factor de crecimiento nervioso era una proteína o un ácido nucleico, Cohen y uno de nosotros, Levi-Montalcini, trató el extracto de células tumorales con veneno de serpiente que contenía una elevada concentración de fosfodiesterasa, enzima que degrada los ácidos nucleicos. Sorprendentemente, observamos que la adición de pequeñas cantidades del veneno de serpiente a la fracción activa de las células del sarcoma-180 incrementaba el efecto promotor del crecimiento de la fracción en lugar de reducirlo. Fue fácil demostrar que era el propio veneno de serpiente la fuente del incremento de actividad. La adición de una pequeña cantidad de veneno al medio de cultivo en ausencia del extracto de sarcoma-180 determinó el crecimiento del mismo halo denso de fibras nerviosas alrededor del ganglio sensitivo o simpático aislados.

El factor de crecimiento nervioso del veneno de serpiente estaba mucho más concentrado y era más potente que el de las células del sarcoma-180. Cohen pudo purificar el factor del veneno y demostró que se trataba de una proteína. Las inyecciones de FCN, purificado a partir de veneno de serpiente, en embriones de pollo determinaron un exceso de inervación simpática de vísceras y vasos sanguíneos similar al inducido por células de sarcoma-180.

El hallazgo de dos fuentes de FCN no relacionadas –el sarcoma de ratón y el veneno de serpiente– sugirió la posibilidad de que también otros tejidos podrían segregar el factor. La investigación se centró en la glándula submaxilar de roedores, similar en ciertos aspectos a la glándula venenosa de las serpientes. Cohen aisló un FCN de las glándulas salivales del ratón que era unas 10.000 veces más activo que el purificado del sarcoma-180, y unas 10 veces más ac-



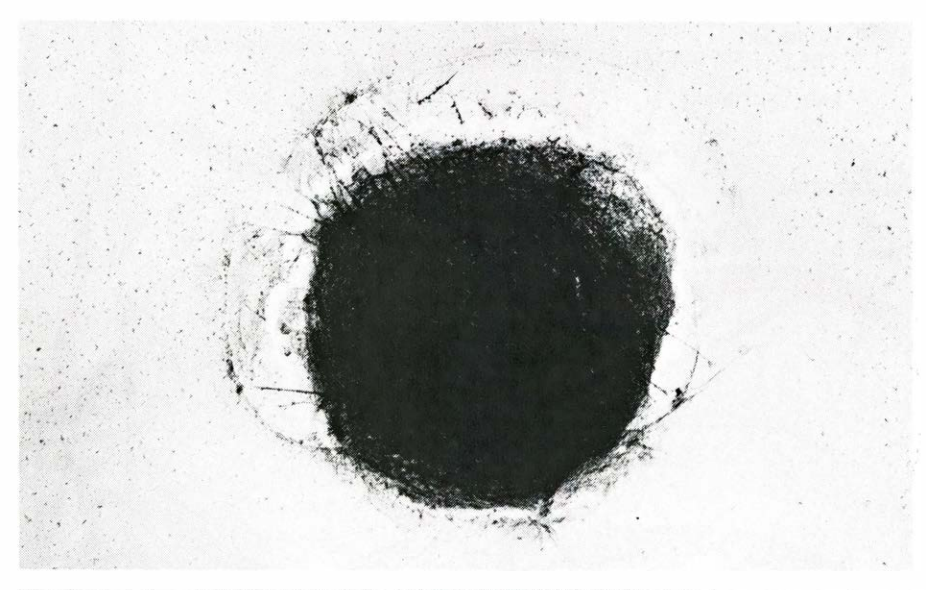
CADENA GANGLIONAR flanqueando la médula espinal de los mamíferos (en este caso se muestra la correspondiente a un ser humano). Los ganglios contienen los cuerpos celulares de las neuronas simpáticas que inervan los distintos órganos, vasos sanguíneos, musculatura lisa y glándulas del organismo. Los nódulos de forma irregular se mantienen unidos mediante fibras nerviosas y están conectados con las raíces de la médula espinal. Cada ganglio inerva los vasos sanguíneos y los órganos del segmento correspondiente del tronco y extremidades. Durante el desarrollo del sistema nervioso, las neuronas de los ganglios emiten fibras para la inervación de los órganos efectores específicos. Puesto que los ganglios son muy accesibles, pueden ser estudiados con facilidad. (Ilustración de Carol Donner.)

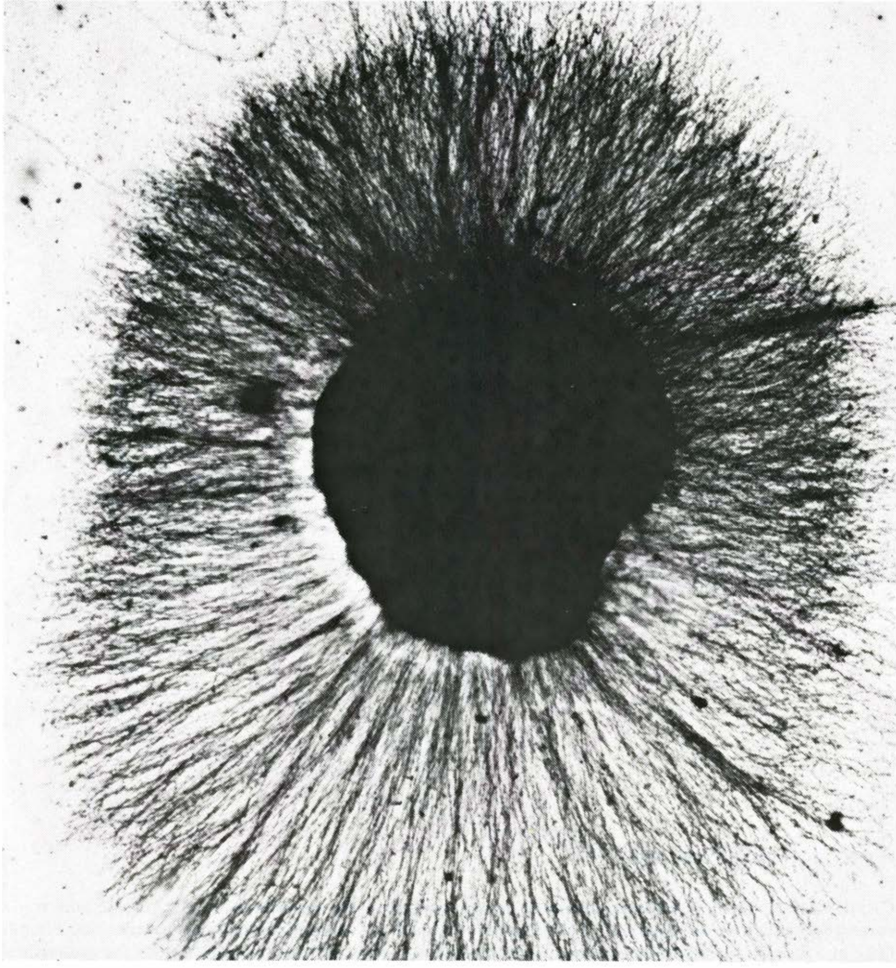
tivo que el purificado a partir del veneno de serpiente. En las dos décadas siguientes se demostró que pequeñas cantidades de FCN eran segregadas por una gran variedad de células normales y neoplásicas (cancerosas).

En 1969, V. Bocchini y Pietro U. Angeletti, del laboratorio de Biología Celular de Roma, idearon un método para purificar el FCN de las glándulas salivales del ratón. Partiendo de notables cantidades de FCN obtenidas utilizando esta técnica, Ruth Hogue Angeletti y Ralph A. Bradshaw, de la Facultad de Medicina de la Universidad de Washington, pudieron determinar la secuencia de aminoácidos de la proteína. El FCN biológicamente activo es un dímero, es decir, un complejo de dos cadenas polipeptídicas idénticas, cada una con un peso molecular de 13.250 dalton. La molécula de FCN tiene un marcado predominio de aminoácidos cargados positivamente y presenta una carga positiva neta a *p*H neutro. El plegamiento de la cadena polipeptídica está estabilizado por tres puentes covalentes disulfuro establecidos entre cisteínas que ocupan diferentes posiciones a lo largo de la cadena. La presencia de puentes disulfuro es común en proteínas que, como la insulina y los anticuerpos, son segregadas activamente por las células. Las restricciones que tales puentes imponen a la conformación de la cadena polipeptídica son aparentemente necesarias para evitar su desnaturalización e inactivación en un medio como el extracelular sujeto a condiciones mucho más adversas que el interior de la célula.

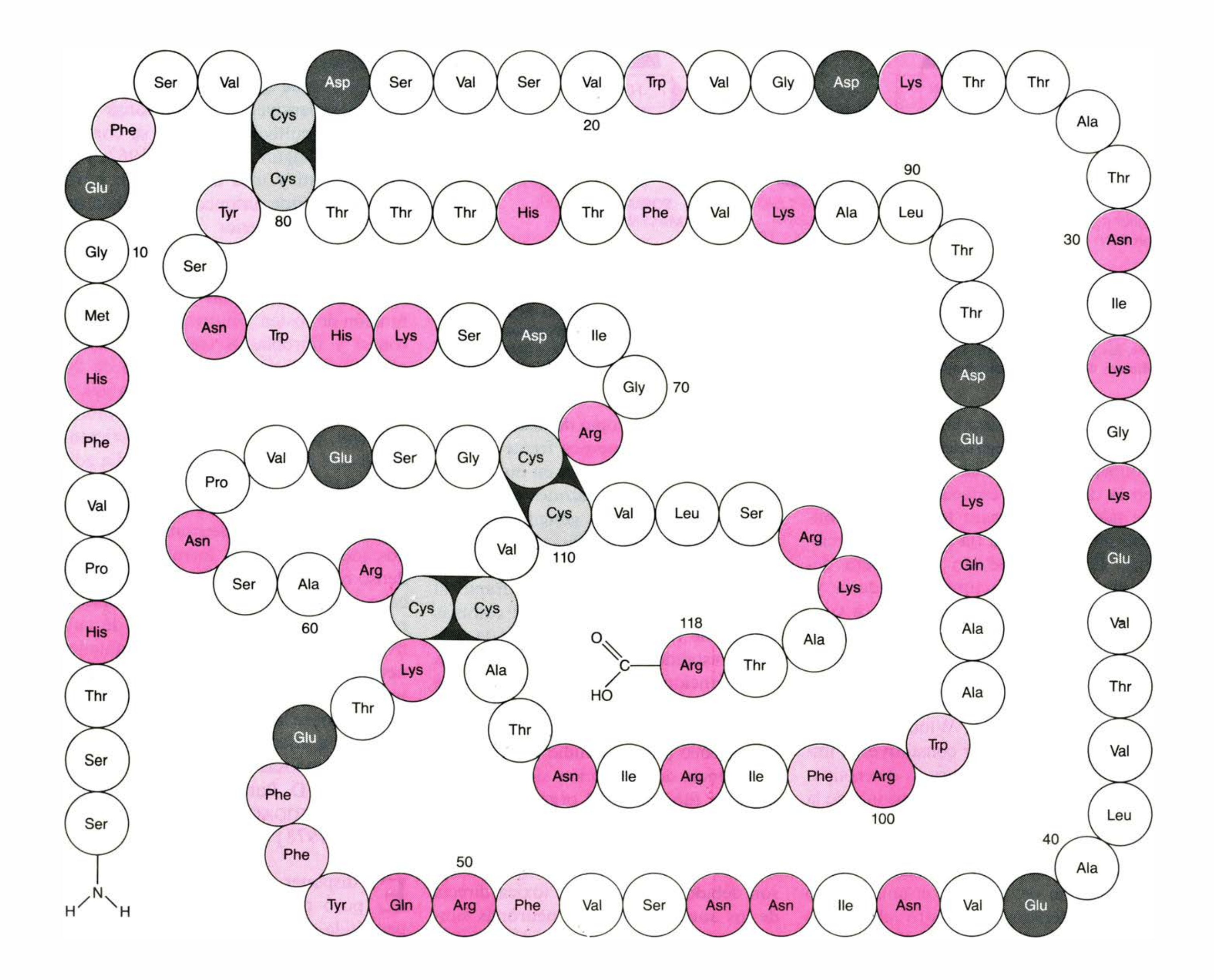
La comparación de la secuencia de aminoácidos del FCN y de otros polipéptidos, llevada a cabo por William A. Frazier, de la Universidad de Washington, reveló que el FCN y la insulina poseen ciertas secuencias comunes. Esta observación condujo a la hipótesis de que el gen para el FCN evolucionó a partir de un gen ancestral para la proinsulina, el precursor polipeptídico de gran tamaño que al ser hidrolizado da lugar a la molécula activa de insulina. Frazier sugirió que el gen ancestral de la proinsulina se habría duplicado y que las dos copias habrían evolucionado de forma divergente dando lugar a la proinsulina y a un polipéptido precursor del FCN. Aunque tal intrigante posibilidad sigue siendo teórica, el parecido de las secuencias de aminoácidos del FCN y de la insulina no es lo suficientemente grande para que dé lugar a una actividad funcional semejante: las dos moléculas tienen células efectoras y actividades biológicas completamente distintas.

En 1967, Silvio S. Varon y Eric M. Shooter, de la Facultad de Medicina de la Universidad de Stanford, aislaron el FCN por un procedimiento distinto y observaron que el dímero del FCN formaba un complejo estable con otras dos proteínas, las llamadas subunidades alfa y gamma. El complejo constaba de dos copias de cada una de estas subunidades y del dímero del FCN. Esta observación resultaba difícil de interpretar, puesto que el FCN por sí solo era biológicamente activo. Las investigacio-





ENSAYO BIOLOGICO DEL FCN basado en la propiedad del factor de inducir la emisión de fibras nerviosas en los ganglios aislados. Las micrografías muestran ganglios sensitivos aislados de un embrión de pollo de ocho días y cultivado durante 12 horas en ausencia (*arriba*) o en presencia (*abajo*) de 10 microgramos de factor de crecimiento nervioso. Las fibras fueron teñidas después de la incubación.



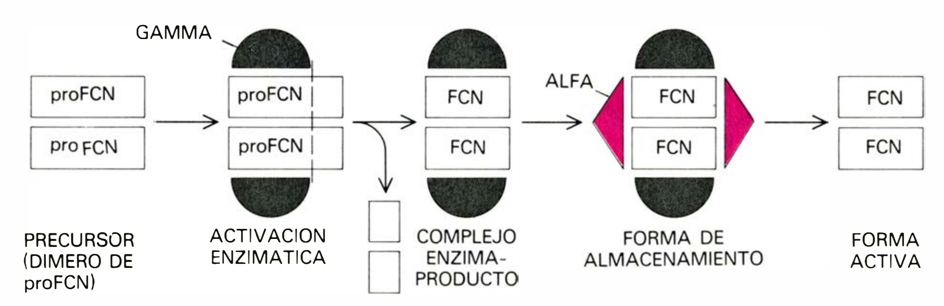
Ala	ALANINA	Leu	LEUCINA
Arg	ARGININA	Lys	LISINA
Asn	ASPARAGINA	Met	METIONINA
Asp	ACIDO ASPARTICO	Phe	FENILALANINA
Cys	CISTEINA	Pro	PROLINA
Gln	GLUTAMINA	Ser	SERINA
Glu	ACIDO GLUTAMICO	Thr	TREONINA
Gly	GLICINA	Trp	TRIPTOFANO
His	HISTIDINA	Tyr	TIROSINA
Ile	ISOLEUCINA	Val	VALINA

ESTRUCTURA PRIMARIA DEL FCN representada esquemáticamente como una cadena de aminoácidos. Tres puentes covalentes unen los grupos sulfhidrilo de las cisteínas. Al igual que en otras proteínas, la secuencia de aminoácidos da lugar a una estructura tridimensional específica que se origina como consecuencia de una serie de interacciones débiles de los aminoácidos. El patrón de plegamiento específico del FCN no ha sido aún determinado. Las distintas clases de aminoácidos se han representado de la forma siguiente: los de carga positiva, *color oscuro*; los de carga negativa, *gris oscuro*; los que poseen un anillo aromático, *color claro*; los que contienen átomos de azufre, *gris claro*; y los aminoácidos sin carga, *blanco*. Por encima de una concentración crítica las dos cadenas monoméricas de FCN, cada una con un peso molecular de 13.250 dalton, se asocian para formar un dímero. No se sabe todavía si es el monómero o el dímero la forma molecular de FCN biológicamente activa.

nes ulteriores revelaron que la subunidad gamma era un enzima específico que hidrolizaba las cadenas polipeptídicas en un punto adyacente al aminoácido arginina. La subunidad alfa, por otra parte, no tenía actividad biológica detectable. La extraña asociación de estas tres proteínas dispares reclamaba una explicación.

La hipótesis de Shooter atribuía a las subunidades alfa y gamma las funciones de activación, almacenamiento y protección de la molécula de FCN. De acuerdo con su esquema, el producto inicial del gen en la síntesis del FCN es un precursor polipeptídico de gran tamaño llamado proFCN, dos copias del cual forman un dímero. Las dos subunidades gamma se unen entonces al dímero proFCN y lo hidrolizan dando lugar al dímero FCN activo. A diferencia del complejo típico que forman el enzima y su producto, que se disocia rápidamente, las subunidades gamma permanecen unidas al dímero FCN después que la hidrólisis de las cadenas de proFCN ha tenido lugar. (Las dos subunidades alfa podrían unirse al dímero proFCN antes de su hidrólisis, o al complejo de las subunidades gamma y el dímero FCN después de la hidrólisis.) La asociación de las subunidades alfa y gamma con el dímero FCN sirve aparentemente para protegerlo de la degradación ulterior por enzimas proteolíticas presentes en los líquidos biológicos. La hipótesis de Shooter recibió apoyo experimental recientemente cuando Edward A. Berger aisló la molécula de proFCN y demostró que la incubación de este precursor con la subunidad gamma daba lugar a su completa conversión en FCN activo.

La actividad biológica de una proteína depende de su estructura tridimensional, la cual consiste no sólo en la secuencia



SINTESIS DE FCN mostrando distintas etapas de "maduración". La forma inicial de la molécula es una cadena precursora denominada proFCN que forma un dímero. Las cadenas de proFCN son hidrolizadas por un enzima conocido como subunidad gamma, dando lugar al dímero FCN. Las dos subunidades gamma permanecen asociadas al dímero después de la hidrólisis y dos proteínas adicionales, las subunidades alfa, se unen al complejo. La asociación de las subunidades alfa y gamma con el dímero de FCN aparentemente protege a éste de una ulterior degradación por acción de otros enzimas.

de aminoácidos, sino también en el patrón preciso de plegamiento de la cadena polipeptídica (la estructura secundaria) y en la disposición espacial de dos o más de tales cadenas para formar una entidad globular única (la estructura terciaria). El único método fiable para determinar las estructuras secundaria y terciaria de una molécula proteica consiste en cristalizarla y en analizar matemáticamente el patrón de difracción de rayos X del cristal. La reciente cristalización del factor de crecimiento nervioso llevada a cabo por A. Wlodawer, Keith O. Hodgson y Shooter constituye un importante paso inicial para la determinación de la estructura tridimensional de la molécula por difracción de rayos X.

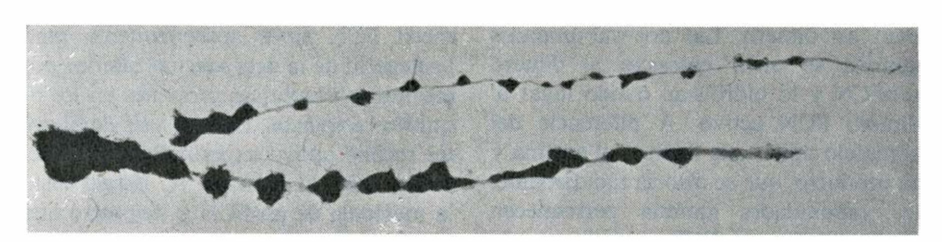
Los primeros datos reveladores de la función del FCN en el animal vivo fueron obtenidos en nuestro laboratorio de la Universidad de Washington en 1959. Cohen obtuvo anticuerpos específicos frente al FCN inyectando en conejos factor de crecimiento purificado a partir de ratones. Pequeñas cantidades de suero de conejo con anticuerpos frente al FCN fueron inyectadas a ratones recién nacidos. Al cabo de un mes, los ratones inyectados y otros control fueron sacrificados. En los ratones tratados con anticuerpos frente al FCN los ganglios simpáticos tenían un tamaño tan reducido que eran difícilmente visibles con el microscopio de disección. Sin embargo, los anticuerpos frente al FCN no producían alteraciones en otros órganos y tejidos, y, por razones desconocidas, no dañaban a los ganglios simpáticos de localización periférica que controlan los órganos sexuales en ambos sexos. Utilizando anticuerpos frente al FCN ha sido posible producir colonias enteras de ratones que alcanzan la madurez estando totalmente desprovistas de la función nerviosa simpática y que son normales en todos los demás aspectos. Estos animales constituyen un modelo idóneo para el estudio del efecto de la ausencia de inervación sim pática en distintas funciones del organismo.

No se conoce todavía si los dramáticos efectos de carácter selectivo producidos por los anticuerpos frente al FCN son debidos a una acción tóxica directa de los anticuerpos sobre neuronas simpáticas inmaduras o a una inactivación del FCN circulante producida por los anticuerpos frente al mismo. En este segundo caso el efecto sería indirecto y la muerte de las neuronas simpáticas se produciría como consecuencia de la deprivación del FCN, factor que requieren para sobrevivir. Esta segunda alternativa parece la más probable. La obtención de pruebas convincentes en su favor proporcionaría la evidencia de que el factor de crecimiento nervioso es un requerimiento absoluto para la supervivencia y crecimiento de las neuronas simpáticas inmaduras en el animal vivo.

El hecho de que el FCN es necesario para la supervivencia de neuronas simpáticas en cultivo de tejidos se conoce desde hace algún tiempo. En 1963, Pietro Angeletti y uno de nosotros, Levi-Montalcini, que entonces trabajaban en el Instituto Superior de Sanidad de Roma, disecaron ganglios sensitivos y simpáticos en sus componentes celulares: neuronas, células de la glía (con función de sostén y nutrición de las neuronas) y fibroblastos (células de tejido conectivo embrionario). Al cabo de 24 horas de cultivo, las células de la glía y los fibroblastos sobrevivían y se multiplicaban, pero las neuronas experimentaban una degeneración masiva. La adición diaria de pequeñas cantidades de FCN al medio de cultivo determinaba la supervivencia de las neuronas por períodos indefinidos y la formación de un denso retículo de fibras nerviosas que al cabo de pocos días cubría completamente la superficie de la placa de cultivo. La propiedad del FCN de mantener neuronas simpáticas en cultivo tisular ha permitido realizar ingeniosos experimentos sobre la diferenciación neuronal [véase "La diferenciación química de las células nerviosas", por Paul H. Patterson, David D. Potter y Edwin J. Furshpan; INVESTIGACION Y CIENCIA, septiembre, 1978]

El disponer de cantidades de FCN puro del orden de miligramos ha permitido ensayar sus efectos en el organismo vivo. Todos los mamíferos responden al FCN de la misma forma, aunque por razones prácticas los roedores son los animales de elección para los experimentos. Al inyectar 10 microgramos de FCN por gramo de peso corporal en roedores recién nacidos durante períodos de hasta tres semanas, sus ganglios simpáticos experimentaban un aumento de tamaño y llegaban a ser de 10 a 12 veces mayores que los de los animales de control. Este incremento excesivo del tamaño de los ganglios simpáticos podría estar en relación con tres procesos distintos: 1) un incremento de la velocidad de diferenciación de las neuronas simpáticas, 2) un aumento del número total de neuronas en el ganglio, y 3) un incremento en el volumen de las neuronas completamente diferenciadas.

El FCN no incrementa el número de neuronas en los ganglios simpáticos estimulando su multiplicación. Tal como propuso en primer lugar I. A. Hendry, el marcado incremento en el número de neuronas obedece a una mayor supervi-



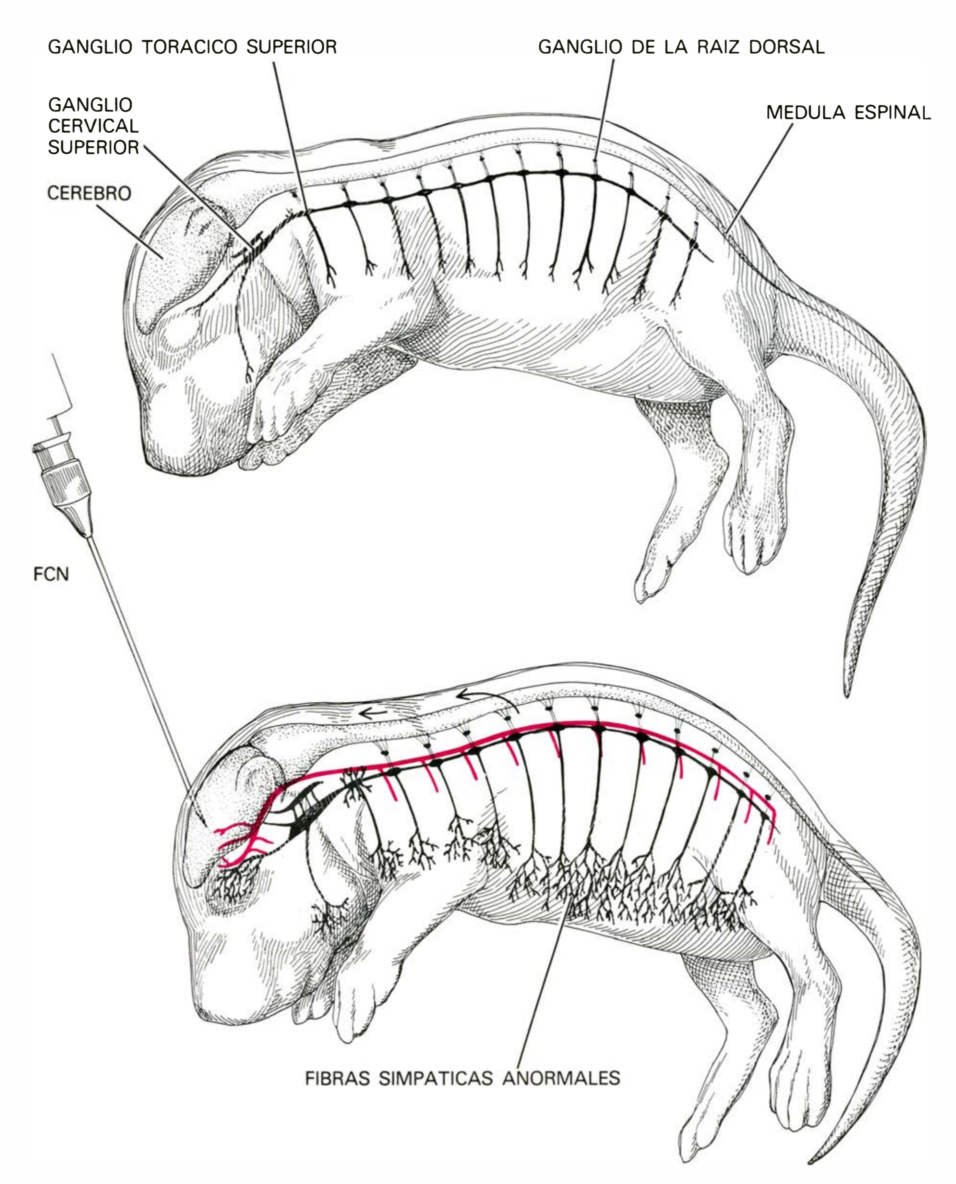
EL FCN INCREMENTA EL CRECIMIENTO de la cadena ganglionar simpática. Las dos series de ganglios fueron aisladas de ratones de tres semanas de edad y procedentes de la misma camada. El ratón control (*arriba*) fue inyectado diariamente con una disolución salina fisiológica, mientras que el otro ratón (*abajo*) recibió 10 microgramos de factor de crecimiento nervioso (FCN) por gramo de peso corporal. La cadena ganglionar mostró, como resultado final, un notable incremento de volumen.

vencia de neuronas inmaduras superfluas que, ordinariamente, morirían en el curso del desarrollo. La muerte celular es un acontecimiento común en la formación del sistema nervioso: poblaciones enteras de neuronas inmaduras mueren o su número se ve reducido drásticamente. De hecho, en los primeros estadios del desarrollo del embrión de pollo, el número de células muertas en los ganglios sensitivos y simpáticos sobrepasa a menudo al número de células vivas. La hipótesis aceptada generalmente supone que las neuronas inmaduras que no establecen conexiones funcionales con sus células efectoras están condenadas a morir; la existencia de neuronas superfluas es una garantía de que se establecerán las conexiones apropiadas. El FCN administrado experimentalmente determina que las neuronas superfluas sobrevivan y se diferencien a pesar de su fracaso en establecer conexiones. A través de este mecanismo, el FCN incrementa sustancialmente el número de neuronas en los ganglios simpáticos.

El FCN, además de ser aparentemente esencial para la supervivencia de las neuronas simpáticas inmaduras, desempeñaría un papel vital dirigiendo las fibras nerviosas hacia sus correspondientes órganos efectores. Se han propuesto tres mecanismos básicos para explicar la formación de circuitos neuronales específicos: 1) un programa genéticamente predeterminado en cada neurona, que se expresa según reglas rígidas y no modificables; 2) un proceso al azar de prueba y error en el que las fibras nerviosas en crecimiento que establecen conexiones adecuadas quedan consolidadas y aquellas que fracasan son reabsorbidas; y 3) un programa general de formación de circuitos que dependería de la interacción de factores genéticos y factores extrínsecos.

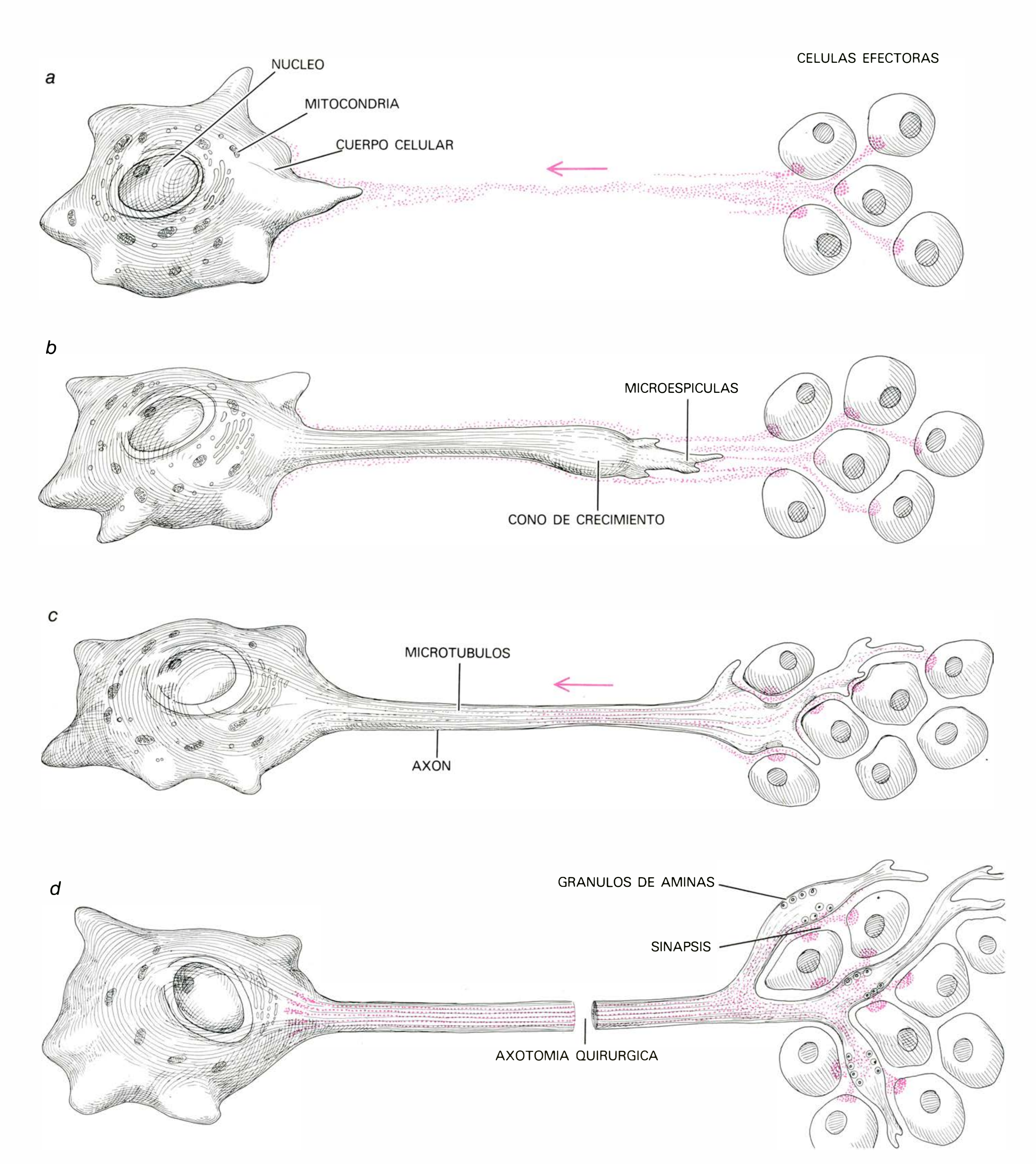
El primer mecanismo puede descartarse, pues requeriría una enorme cantidad de información genética, mucho mayor que la que puede codificarse en el ADN del núcleo de cada neurona. El segundo mecanismo es también improbable, puesto que la formación al azar de los circuitos neuronales sería extraordinariamente lenta y supondría un derroche de energía y materias primas. El tercer mecanismo parece, por tanto, el más probable: los circuitos neuronales se establecerían a través de alguna combinación de factores genéticos y factores extrínsecos.

La existencia de tales factores extrínsecos fue sugerida por primera vez por Cajal, que los concibió como señales químicas que emanaban de los tejidos periféricos dirigiendo el crecimiento de las fibras nerviosas hacia sus células efectoras. Cajal denominó a este proceso neurotropismo. Su hipótesis fue olvidada durante muchos años al no disponerse de metodología adecuada para detectar tales factores químicos en el embrión vivo. El descubrimiento y la obtención del FCN proporcionó una oportunidad para reconsiderar el concepto de neurotropismo en condiciones experimentales más favorables.



EFECTO NEUROTROPICO DEL FCN demostrado por inyección del factor durante 10 días en el cerebro de ratas recién nacidas; los animales control fueron inyectados sólo con disolución salina fisiológica. En los animales tratados con FCN (*abajo*), el factor de crecimiento difundió desde el sitio de inyección a la médula espinal y alcanzó los ganglios simpáticos adyacentes a través de las raíces motoras y sensitivas. El FCN indujo el crecimiento de fibras nerviosas simpáticas (*coloreadas*) que alcanzaron la médula espinal a través de la raíces espinales y se extendieron hasta alcanzar el punto de inyección en el cerebro. Esta proyección anormal de fibras simpáticas (que inervan normalmente tejidos periféricos) en respuesta al FCN apoya la hipótesis de que un gradiente de FCN dirige las fibras simpáticas en crecimiento en la periferia. Por otra parte, la inyección de FCN en el cerebro o en la circulación general aumentaba las ramificaciones periféricas de las fibras simpáticas. Los experimentos fueron realizados por M. L. Menesini Chen y J. S. Chen en el laboratorio de los autores en Roma.

La demostración del papel del FCN en la formación de los circuitos neuronales del sistema nervioso simpático se basa en experimentos realizados en el animal y en el tubo de ensayo. En el laboratorio de Biología Celular de Roma, se inyectó FCN en el cerebro de roedores recién nacidos. Un efecto inesperado de este tratamiento fue el hecho de que las fibras nerviosas surgidas de los ganglios simpáticos invadían el cerebro y la médula espinal. Aparentemente, el FCN inyectado en el cerebro difundía a través de las raíces sensitivas y moto-



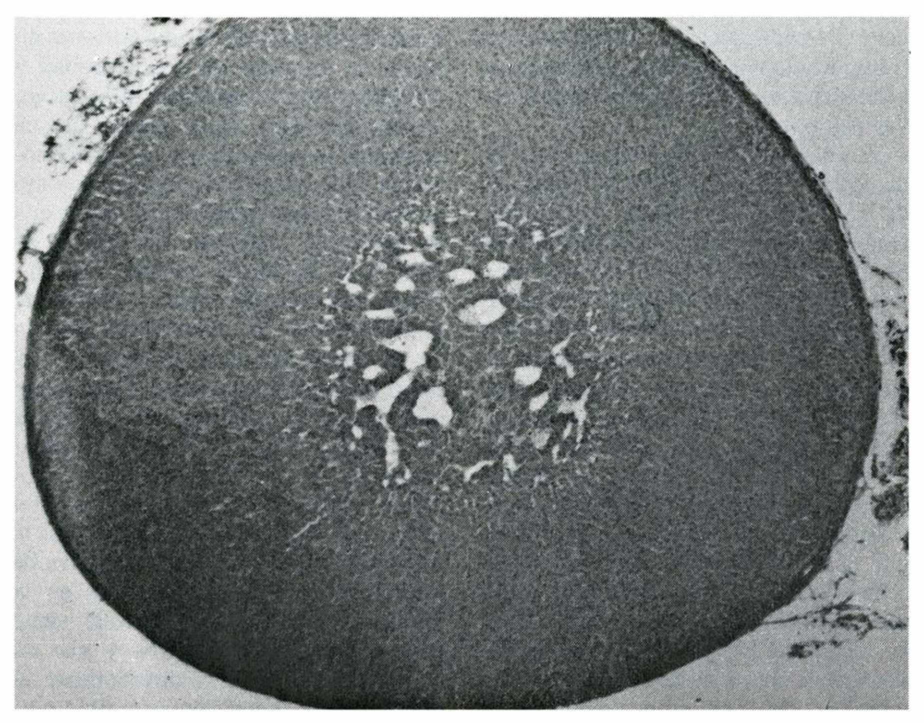
ELONGACION DE LA FIBRA NERVIOSA de una neurona simpática inmadura inducida y dirigida por el FCN. En condiciones normales, las células efectoras inervadas por las neuronas simpáticas o las que se hallan presentes a lo largo de la vía de crecimiento del axón, sintetizan y segregan pequeñas cantidades de FCN. El factor de crecimiento difunde en el espacio intercelular y se une a moléculas receptoras específicas de la superficie de la neurona inmadura (*a*). La interacción del FCN y su receptor desencadena (a través de un mecanismo aún desconocido) la interacción en el interior de la neurona de las proteínas filamentosas llamadas microtúbulos y microfilamentos. Estas estructuras son constituyentes esenciales del cono de crecimiento, el orgánulo locomotor de los axones que experimentan elongación. El cono de crecimiento, situado en el extremo de la fibra nerviosa, posee prolongaciones exploradoras móviles llamadas microespículas, que confieren movimiento y proyectan en una dirección determinada al axón en crecimiento (*b*). Una vez el axón ha establecido contactos definitivos con sus células efectoras, las conexiones se consolidan en uniones físicas y funcionales conocidas como sinapsis. El FCN producido por las células efectoras es conducido directamente a través del axón hacia el cuerpo celular mediante un mecanismo denominado transporte axónico retrógrado, que, probablemente, se debe a la acción de estructuras filamentosas del axón (*c*). El bloqueo del transporte axónico retrógrado (seccionando el axón, destruyendo la sinapsis con 6-hidroxidopamina, o interrumpiendo el flujo axoplásmico con vinblastina) da lugar a la degeneración de las neuronas simpáticas ganglionares (*d*). Todos los tratamientos citados ejercen sus efectos letales evitando el acceso normal del FCN a las neuronas simpáticas desde los tejidos periféricos inervados a través del transporte axónico retrógrado. Los anticuerpos frente al FCN destruyen las neuronas simpáticas inmaduras al inactivar el FCN de modo que ya no puede unirse a las neuronas.

ras de la médula espinal y alcanzaba la cadena ganglionar simpática que flanquea la médula, donde inducía el crecimiento de fibras nerviosas. Esta observación implica que el extremo de una fibra nerviosa simpática en crecimiento se extiende a lo largo de un gradiente de difusión de FCN liberado por el órgano efector. El gradiente de FCN no determina completamente el curso de las fibras nerviosas, pero ayuda a orientarlas en la dirección adecuada.

Los elegantes experimentos llevados a cabo por Robert B. Campenot, de la Facultad de Medicina de Harvard, corroboraron la existencia de un neurotropismo mediado por el FCN. Este investigador ideó un sistema de cultivo consistente en tres compartimientos separados por barreras impermeables de grasa de silicona estéril que impedían la difusión de líquido de un compartimento a otro. Las neuronas simpáticas inmaduras se cultivaron en el compartimento central en presencia de FCN, y sus fibras nerviosas en crecimiento podían alcanzar los compartimentos laterales pasando por debajo de las barreras de silicona a través de grietas practicadas en la base de la placa de cultivo. Campenot llenó uno de los compartimentos laterales con una disolución nutritiva que contenía FCN y el otro compartimento con la misma disolución desprovista de FCN. Observó que las fibras nerviosas se dirigían únicamente al compartimento lateral que contenía FCN. Cuando del compartimento lateral se retiraba la solución rica en FCN, las fibras nerviosas que habían alcanzado dicho compartimento empezaban a degenerar y, sin embargo, los cuerpos neuronales del compartimento central en presencia de FCN permanecían en buenas condiciones.

Estos experimentos resolvieron la vieja cuestión de si los factores químicos extrínsecos realizaban una función importante dirigiendo el crecimiento de las fibras nerviosas. Puesto que el FCN se libera en cantidades muy pequeñas en los tejidos periféricos que reciben inervación de los ganglios simpáticos, parece evidente que un gradiente de difusión de FCN dirige las fibras hacia sus correspondientes células efectoras. Cuando la fibra nerviosa en crecimiento y la célula efectora establecen finalmente contacto, la adhesión provisional se consolida constituyéndose la organización estructural y funcional de la sinapsis.

Una vez que las fibras nerviosas han establecido las conexiones sinápticas adecuadas con las células efectoras, la supervivencia de las células nerviosas en el ganglio parece depender de la disponibilidad de FCN. Investigaciones llevadas a cabo por Hendery, de la Universidad Nacional de Australia, y por K. Stöckel y H. Thoenen, del Instituto de Inmunología de Basilea, demostraron que el FCN es captado por las terminaciones nerviosas de las fibras simpáticas y transportado hacia los cuerpos neuronales a través del axón. El transporte retrógrado del FCN por el axón es absolutamente esencial para la supervivencia de las neuronas. Cuando se corta el transporte (por sección de los axones, por tra-





EL FCN TRANSFORMA LAS CELULAS ADRENALES en neuronas simpáticas al ser inyectado en ratas antes y después del nacimiento. Las secciones ecuatoriales de las glándulas adrenales de ratas de dos días de edad muestran el efecto de la inyección de disolución salina sin FCN (*arriba*) y de FCN (*abajo*). La glándula adrenal normal consiste en una densa capa cortical de células que rodean una zona medular central constituida por células adrenales y vasos sanguíneos. El tratamiento con FCN determina una total reestructuración de la glándula. La médula central aumenta enormemente de volumen como resultado de la transformación de las células adrenales en neuronas simpáticas que irradian un gran número de fibras profusamente ramificadas. Estas fibras forman un gran haz que reemplaza la capa cortical interna y cruza la corteza, disponiéndose circularmente alrededor de la zona externa. Los experimentos fueron realizados por Luigi Aloe, de nuestro laboratorio de Biología Celular de Roma.

tamiento con la droga vinblastina que bloquea el transporte axónico, o por administración de 6-hidroxidopamina, que destruye las terminaciones nerviosas), las neuronas simpáticas ganglionares mueren. Los efectos letales del bloqueo del transporte de FCN pueden contrarrestarse proporcionando a los cuerpos neuronales FCN exógeno. En este caso, el FCN administrado compensa el déficit del FCN endógeno que normalmente es transportado por vía retrógrada desde la célula efectora al cuerpo neuronal a través del axón.

El trabajo de distintos laboratorios ha demostrado que el transporte retrógrado de FCN por el axón sigue a su interacción con sitios receptores específicos de las terminaciones de las fibras nerviosas. Los receptores son proteínas localizadas en la superficie externa de la membrana celular; estas moléculas proporcionan sitios de reconocimiento específico para substancias que transmiten mensajes, tales como hormonas, neurotransmisores y factores de crecimiento. La existencia de tales receptores específicos en la superficie de las neuronas hace posible que el FCN ejerza sus acciones a concentraciones extraordinariamente bajas (del orden de 2,8 microgramos por litro). La unión del FCN a sus receptores desencadena una serie de fenómenos bioquímicos que conducen finalmente al crecimiento de la fibra nerviosa.

Las neuronas simpáticas inmaduras responden al FCN con una explosión de actividad metabólica, que proporciona el material necesario para el crecimiento de la fibra nerviosa y la síntesis de moléculas de neurotransmisor. La célula-sintetiza más proteínas y lípidos, capta aminoácidos del medio extracelular y consume a mayor velocidad glucosa y otros compuestos ricos en energía. Estos efectos pueden considerarse como una respuesta general a una señal primitiva transmitida por el FCN a través de los receptores de la superficie celular.

Poco tiempo después de la unión del FCN con sus receptores, los constituyentes proteicos del citoplasma de la neurona simpática inmadura experimentan una profunda reorganización. Particularmente, las estructuras filamentosas llamadas microtúbulos y microfilamentos llenan todo el espacio disponible entre el núcleo celular y la membrana de la célula. Estos filamentos desempeñan un papel esencial en el crecimiento de la fibra nerviosa proporcionando un armazón estructural y la fuerza propulsiva para la elongación.

¿De qué manera el FCN controla la unión de estas proteínas filamentosas? Una posibilidad es que actúe directamente incrementando la polimerización de la tubulina y actina, las proteínas monoméricas que dan lugar, respectivamente, a los microtúbulos y a los microfilamentos. En el laboratorio de Biología Celular de Roma, ensayamos la hipótesis midiendo la velocidad de unión de la tubulina y de la actina en el tubo de ensayo en presencia y en ausencia de FCN. En ausencia de FCN, las disoluciones diluidas de proteínas no polimerizaban y no daban lugar a filamentos (o lo hacían a velocidad muy baja), puesto que la tendencia termodinámica de los monómeros a permanecer en disolución como moléculas independientes era mayor que su tendencia a agregar. La adición de FCN a la disolución inducía, sin embargo, una reacción rápida y masiva de polimerización. Investigaciones ulteriores revelaron que el FCN reúne el número mínimo de monómeros necesarios para iniciar la polimerización, después de lo cual la reacción prosigue espontáneamente hasta su terminación. Tomando como base estos resultados obtenidos en el tubo de ensayo, sugerimos la hipótesis de que la formación masiva de microtúbulos y microfilamentos en la neurona simpática en desarrollo viene desencadenada por el FCN, y que tal efecto puede conducir directamente al crecimiento y elongación de la fibra nerviosa.

Recientemente se ha descubierto una función del FCN del todo distinta. Ciertas células no nerviosas respondían a la proteína. Lloyd A. Greene y Arthur S. Tischler, de la Facultad de Medicina de Harvard, demostraron en 1977 que las células de la línea designada PC12 (derivadas de un tumor de rata) responden al FCN adquiriendo propiedades características de las neuronas simpáticas. Las células tratadas con FCN emiten fibras, se vuelven eléctricamente excitables y almacenan y liberan neurotransmisores del tipo de las catecolaminas. Cuando se elimina el FCN del medio de cultivo, las fibras se retraen, se pierden las otras propiedades neuronales y aparece de nuevo la proliferación incontrolada característica de las células neoplásicas.

Poco después de este descubrimiento, K. Unsicker y sus colegas de la Universidad Johns Hopkins observaron que las células cromafines obtenidas de la médula adrenal (parte interna de la glándula) y cultivadas en presencia de FCN adquirían las propiedades bioquímicas y morfológicas de las neuronas simpáticas. Experimentos ulteriores llevados a cabo en el laboratorio de Biología Celular de Roma por Luigi Aloe y uno de nosotros, Levi-Montalcini, demostraron que esta extraordinaria transformación puede ocurrir también en el animal vivo. En efecto, inyecciones repetidas de FCN en fetos de rata, y continuadas después del nacimiento durante dos o tres semanas, determinaban la diferenciación de las células cromafines en neuronas simpáticas en la médula adrenal. Estas observaciones indican que el FCN tiene una misión mucho más amplia en el organismo vivo de lo que se había supuesto.

La investigación del FCN, que se inició hace más de dos décadas, está todavía lejos de concluirse. Entre las muchas cuestiones no resueltas podría citarse la razón por la que el FCN se sintetiza y segrega por las glándulas venenosas de las serpientes y por las glándulas salivales de los roedores, aunque ninguna de esas glándulas es esencial para la vida del organismo y de las neuronas simpáticas que dependen del FCN para su supervivencia. La molécula se sintetiza y libera también en pequeñísimas cantidades por una gran variedad de células normales y neoplásicas.

El descubrimiento del FCN ha hecho posible manipular experimentalmente neuronas simpáticas con una considerable facilidad tanto en cultivo de tejidos como en el animal vivo. Con la ayuda del FCN y de los anticuerpos frente al mismo, el neurobiólogo es ahora capaz de incrementar hasta diez veces el número de neuronas simpáticas de un animal, o, por el contrario, eliminarlas íntegramente. Por otra parte, las neuronas que segregan catecolaminas en el cerebro son sorprendentemente similares a las neuronas simpáticas de los ganglios periféricos tanto en sus propiedades morfológicas como bioquímicas. La información obtenida en el estudio de las neuronas del simpático ha resultado útil en el tratamiento de alteraciones de las neuronas del sistema nervioso central que segregan catecolaminas, tales como la enfermedad de Parkinson. A. Björklung, B. Bierre y U. Stenevi, de la Universidad de Lund, obtuvieron hace poco resultados preliminares que señalan que las neuronas del cerebro secretoras de catecolaminas responden al FCN con una ramificación profusa de sus fibras nerviosas. Si tales hallazgos se confirman, se dispondrá de un poderoso instrumento para modular la función de determinados circuitos cerebrales que desempeñan un papel crucial en muchos tipos de conducta.

Control automático por inteligencia distribuida

La microelectrónica permite operar una planta industrial con una gran cantidad de "pastillas" interconectadas, que abarcan desde el control de una sola máquina hasta el control integrado de la planta completa

Stephen Kahne, Irving Lefkowitz y Charles Rose

La aparición del computador de control digital a mediados de los años cincuenta inició una revolución en el control de las plantas industriales, revolución que no ha pasado todavía la etapa de infancia. Pero llegará a presentar, sin lugar a dudas, un gran impacto en la tecnología, en la economía y en la sociedad tan profundo como lo tuvo la revolución industrial. En un comienzo se esperaba que el computador permitiría rebasar ampliamente el escueto objetivo de mantener las variables de proceso en unos valores fijos especificados, y que conduciría hacia otro objetivo más interesante, a saber: determinar la forma en que estas variables cambiarían con el tiempo o con relación a otras variables, de suerte que el rendimiento de todo el sistema fuera optimizado de modo continuo. Pero el problema era que los primeros computadores de control estaban seriamente limitados en cuanto a su velocidad, capacidad de memoria y, sobre todo, complejidad de sus programas. Como resultado de ello, muchos de los intentos iniciales de optimización a través del control por computador, aunque acertadamente concebidos, obtuvieron resultados mucho más pobres de lo previsto. En aquel entonces, lo que en definitiva se buscaba era el control automático de toda una planta a través de un computador central de gran tamaño; éste recibiría sus señales de entrada de múltiples sensores, recopilaría sus lecturas y distribuiría las órdenes adecuadas para controlar las variables de proceso de cualquier clase. El objetivo demostró ser demasiado ambicioso.

Con la aparición del miniordenador y más recientemente del microprocesador (el "ordenador construido en una pastilla") ha cambiado el enfoque total del control por computador de los sistemas industriales. El enfoque actual del control de una planta industrial compleja está basado en la idea de control descentralizado, adquisición distribuida de datos e información distribuida del proceso. La ejecución de estos conceptos exige una jerarquización de los computadores, llevando a cabo cada uno las tareas apropiadas a su rango jerárquico. Hoy parece que esta perspectiva facilitará la consecución del difícil objetivo del control totalmente integrado de una planta.

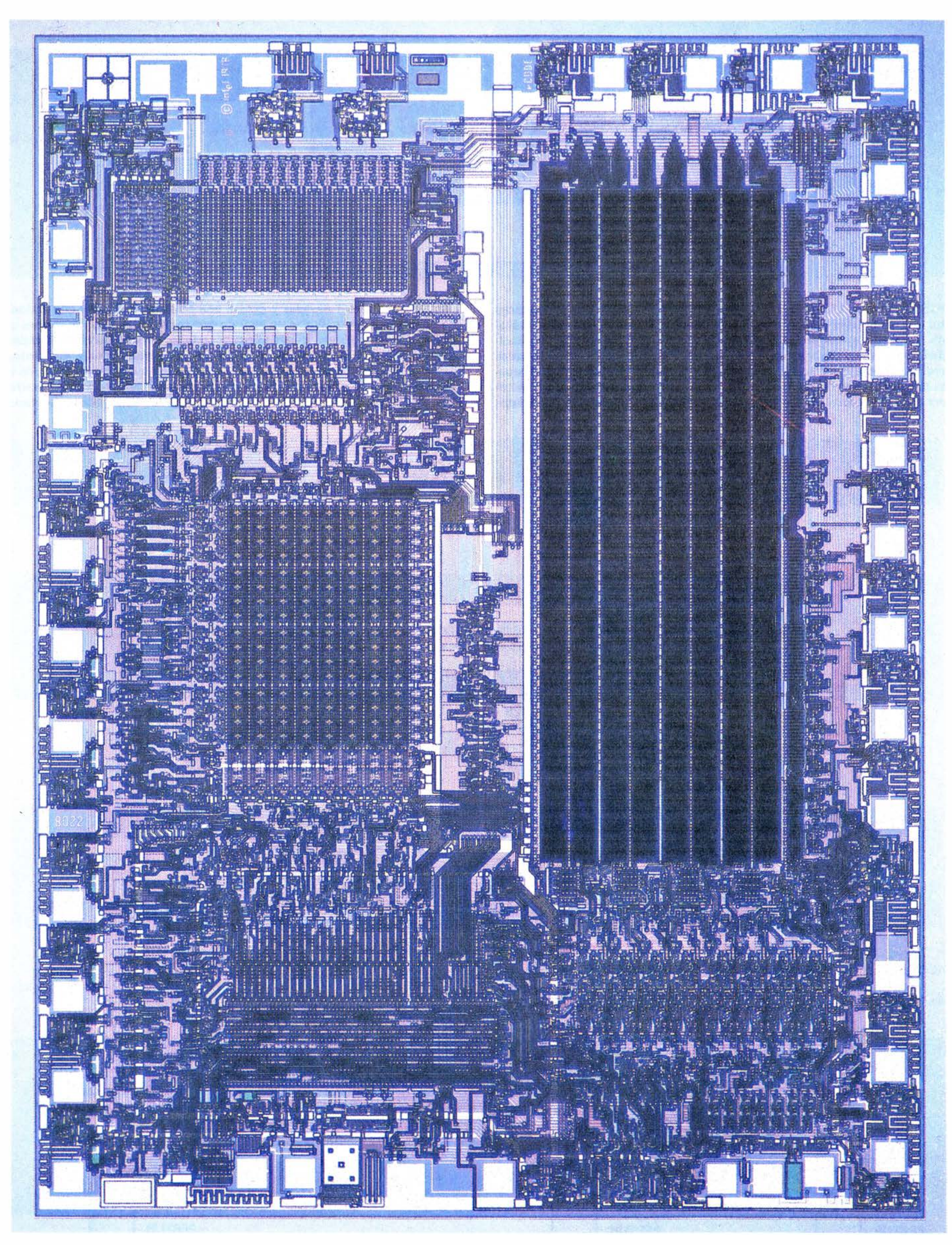
El tratamiento o procesamiento distribuido designa cualquier entorno de cálculo donde múltiples sistemas computadores ejecutan una aplicación dada. El término se emplea comúnmente en el contexto del tratamiento de datos, donde están interconectados múltiples computadores de capacidad media y alta para satisfacer las necesidades de procesamiento de datos de una organización. Si bien el enlace de múltiples microprocesadores, que cooperan estrechamente para realizar el control de un proceso industrial, puede denominarse también tratamiento o procesamiento distribuido, constituye algo cualitativamente diferente. Para describir un entorno de planta controlado por microprocesadores múltiples debemos basarnos, pues, en el término "inteligencia distribuida".

El control automático por inteligencia distribuida tiene, entre otras ventajas, una gran flexibilidad de diseño junto con una fiabilidad y un rendimiento potenciados. Como resultado de ello, la aplicación de los ordenadores a los sistemas de control industrial ha aumentado rápidamente y hoy en día se estima que casi la mitad de las instalaciones de miniordenadores de los Estados Unidos están empeñadas en el control.

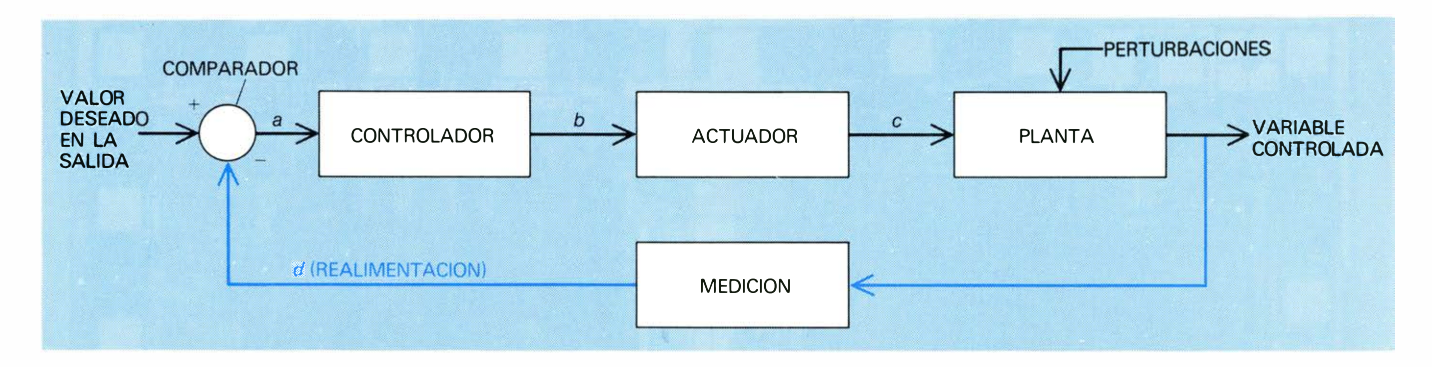
Con los rápidos avances alcanzados paralelamente en el soporte físico (hardware) y en el soporte lógico o conjunto de programas (software) del ordenador y en la teoría de control, podemos proyectar ya sistemas de control en que se consideren todos los factores que influyen en el rendimiento de la planta –interacciones y complejos procesos de realimentación de todas clases– para alcanzar un rendimiento total óptimo. Un sistema integrado avanzado, incluirá, además de las funciones normales de control de procesos, funciones de alto nivel, tales como los procedimientos de arranque y parada, los programas de diagnóstico para responder a acontecimientos imprevistos, sistemas de monitorización y de alarma y aparatos que hagan posible una interacción rápida y eficiente entre el hombre y la máquina.

Usaremos el término control en su sentido más amplio, que abarque todos los aspectos del tratamiento de la información y de la toma de decisiones que den lugar a las acciones que afectan al sistema controlado en una forma deseada. Sin embargo, el objeto del control es aquí la planta industrial. La planta puede ser un equipo productor de bienes o servicios o ser un complejo de unidades de producción que lleguen a formar una compañía completa.

Los factores referentes al funcionamiento del sistema de la planta se consideran entradas. Distinguiremos dos tipos de entradas: entradas de perturbación y entradas de control. Las entradas de perturbación penetran en el sistema a través de fuentes externas y tienden a interferir el comportamiento deseado del sistema. Representan usualmente los efectos de las interacciones con otras unidades de planta y con el medio exterior. Las entradas de control son las generadas por el controlador o por el tomador de decisiones. Suelen derivar de las mediciones realizadas en las salidas



MICROCOMPUTADOR MONOLITICO, Modelo 8022 introducido recientemente por la Intel Corporation. Pertenece a la nueva generación de computadores contenidos en una pastilla que están acelerando la tendencia hacia el control jerarquizado y distribuido de los procesos industriales. El modelo 8022 que mide 0,41 por 0,54 centímetros se venderá en serie por menos de 10 dólares. El microcomputador de ocho bits (ocho dígitos binarios) incluye una memoria de sólo lectura de 16.000 bits (ocupando la mayor parte de la mitad derecha de la pastilla) y una memoria de libre acceso de 512 bits (el rectángulo mayor en la parte media de la mitad izquierda de la pastilla). Para facilitar la interfase con las señales analógicas el aparato incluye un convertidor analógico-digital, conjuntamente con sistemas de interrupción para manejar los sucesos no programados que se presentan con frecuencia en el funcionamiento de plantas industriales. La unidad de procesamiento central de la pastilla, que puede realizar 100.000 operaciones aritméticas y lógicas por segundo, se aloja en la pequeña área de la parte inferior derecha. (Véase *Investigación y Ciencia*, noviembre 1977.)



SISTEMA DE CONTROL DE REALIMENTACION, que proporciona un método general para mantener una variable física en algún valor deseado. La variable a controlar puede ser la temperatura, la presión, la velocidad, el espesor o cualquier otra que pueda cuantificarse. La señal de salida real es medida y convertida en una señal (*d*), que es realimentada y comparada con el punto de consigna o el valor deseado para generar una señal de error (*a*). El controlador es un procesador de datos que analiza las señales de error presentes y pasadas y genera una señal de control (*b*) que acciona un actuador. El actuador es un aparato que convierte la señal de control en una acción en la planta (*c*) que altera la variable, controlada en la dirección deseada. La planta es influida también por las perturbaciones que entran desde el exterior del lazo de control y que tienden a desviar la salida de la planta del valor deseado. Este valor es seleccionado fuera del lazo de control por un operador humano o bien por un controlador automático supervisor.

LIMITACIONES Y OBJETIVOS DE ALTO NIVEL
LIMITACIONES Y OBJETIVOS LOCALES
INFORMACION
SEGUNDO NIVEL
COMPUTADOR COORDINADOR
LIMITACIONES Y OBJETIVOS LOCALES
INFORMACION
LIMITACIONES Y OBJETIVOS LOCALES
INFORMACION
LIMITACIONES Y OBJETIVOS LOCALES
INFORMACION
PRIMER NIVEL
COMPUTADOR LOCAL 1
COMPUTADOR LOCAL 2
COMPUTADOR LOCAL *N*
ACCIONES DE CONTROL
INFORMACION
ACCIONES DE CONTROL
INFORMACION
ACCIONES DE CONTROL
INFORMACION
NIVEL DE PLANTA
COMPUTADOR DE SUBSISTEMA 1
VARIABLES DE INTERACCION
COMPUTADOR DE SUBSISTEMA 2
VARIABLES DE INTERACCION
VARIABLES DE INTERACCION
COMPUTADOR DE SUBSISTEMA *N*
PERTURBACIONES DE ENTRADA
VARIABLES DE SALIDA
PERTURBACIONES DE ENTRADA
VARIABLES DE SALIDA
PERTURBACIONES DE ENTRADA
VARIABLES DE SALIDA

DISTRIBUCION DE LAS FUNCIONES DE CONTROL. Dicha distribución puede obtenerse dividiendo una planta en un número (*N*) de subsistemas y controlando cada uno mediante un computador local, que representa el primer nivel de control. Los controladores locales son responsables del mantenimiento de las variables de los subsistemas a valores coherentes con los objetivos y las limitaciones locales. Los subsistemas interaccionan generalmente entre sí; son necesarios para coordinar las acciones de los controladores locales, de suerte que se cumplan los objetivos de planta.

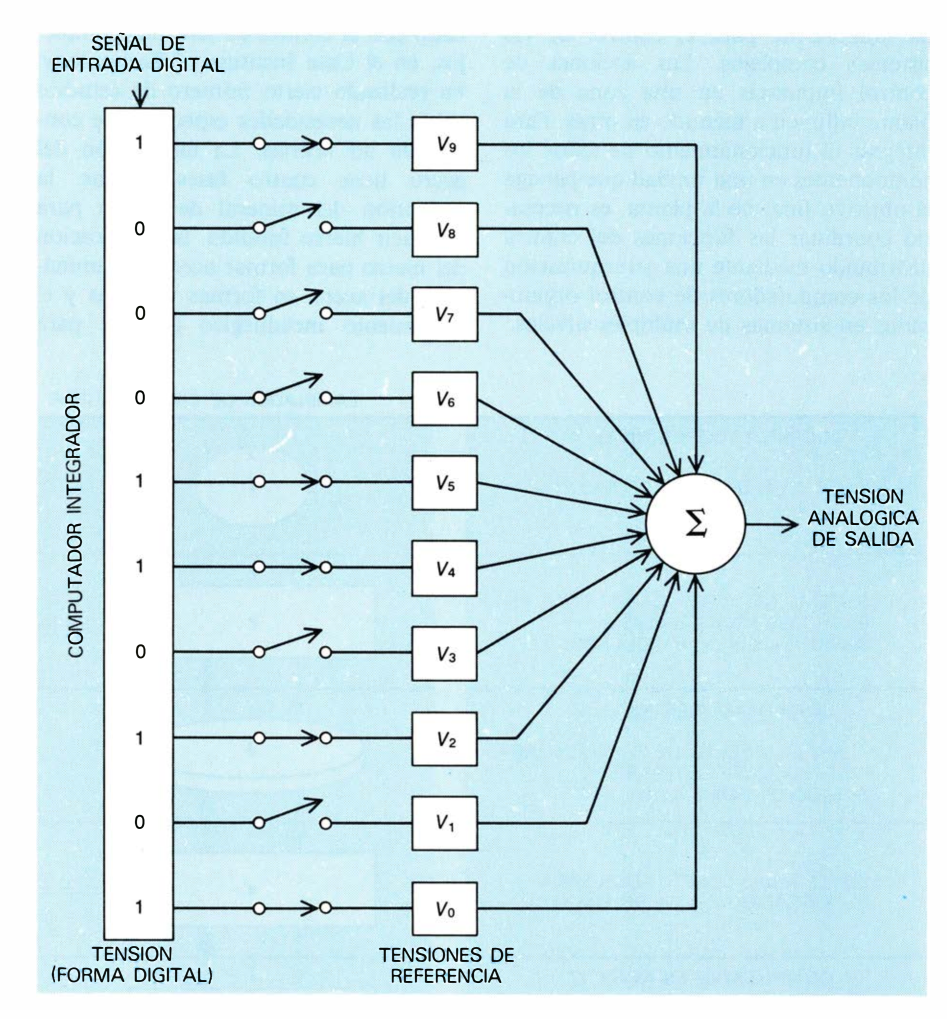
de la planta y están proyectadas para lograr que dichas salidas tomen los valores deseados.

En el funcionamiento de la planta, las salidas se definen como variables que resultan de las entradas y se relacionan con los objetivos totales del sistema. Los objetivos suelen ir más allá del simple mantenimiento de la producción de la planta a un alto nivel constante. Por ejemplo, una acería puede tener asignada la tarea de conseguir una productividad máxima ligada a una contaminación y a un consumo de energía mínimos. El objetivo debe lograrse además dentro de un conjunto de límites permisibles, verbigracia, los impuestos por la capacidad del equipo, los niveles salariales y las normas gubernamentales.

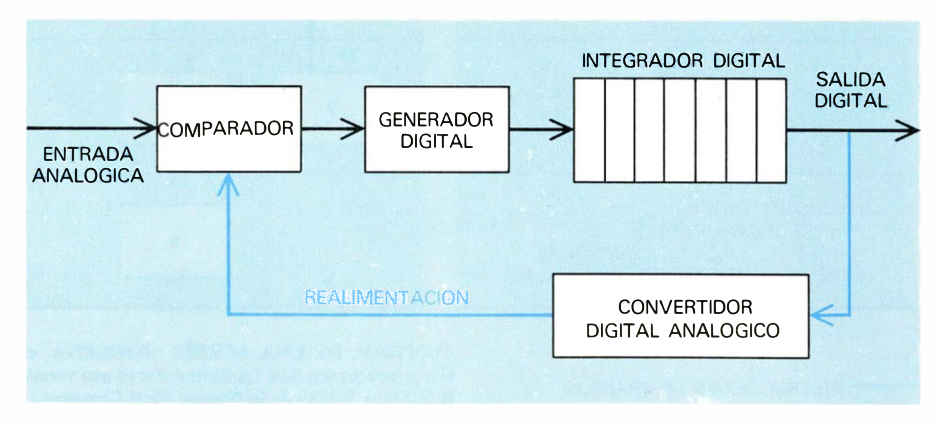
El mayor reto en el control de un sistema quizá sea la definición del objetivo en una forma matemática que represente adecuadamente el comportamiento deseado del sistema y que pueda evaluarse a partir de cantidades medibles en el mismo. La tarea es particularmente difícil en grandes sistemas donde hay que contar no sólo con las variables tecnológicas usuales sino también con las sociales.

A medida que la planta o el sistema a controlar crece en tamaño, el problema del control aumenta rápidamente en complejidad. Las grandes compañías poseen generalmente fábricas en varios lugares y los medios de producción de muchas compañías multinacionales están repartidos por todo el mundo. Muchos equipos están formados por múltiples unidades de producción, cada una de las cuales encierra una gran cantidad de componentes que exigen, a su vez, cientos de elementos de control. Como es natural, la distribución geográfica de estos equipos implica la distribución de otras actividades de la corporación, actividades empresariales y financieras incluidas. Es razonable, pues, distribuir las funciones de captación de datos, de procesamiento de la información y de control, en una forma tal que satisfaga las necesidades de distribución de los equipos de la planta y sus funciones. Al estar la función de control instrumentada en alto grado por el soporte físico digital conviene también que dicho soporte se halle distribuido.

Una idea central en el tema del control es la realimentación. Las señales de realimentación comunican con el elemento de control para indicar de qué modo la variable controlada se desvía del valor o punto de consigna deseado. El elemento de control responde intentando eliminar la desviación. Los lazos de realimentación están incorporados en los sistemas de control distribuido tanto local como globalmente, es decir, en una red amplia. Los controladores de realimentación locales sirven para mejorar la respuesta dinámica y para mantener las variables de la planta en los valores deseados, pese a la existencia de perturbaciones en la entrada. Consideradas globalmente, las informaciones de la realimentación procedentes de varias partes de una gran planta sirven para modificar las políticas productivas y



CONVERSION DE SEÑALES DIGITALES A ANALOGICAS, requerida comúnmente en los sistemas de control industriales en el punto donde una señal controlada digital de bajo nivel necesita ser convertida en una señal de alta potencia para accionar un motor o alguna otra clase de actuador de la planta. La señal digital es convertida en forma analógica asociando cada bit de la señal digital con una tensión de referencia. Si las tensiones de referencia aumentan en un factor de dos desde un nivel al siguiente, V_q tiene 512 veces el valor de V_o. Las tensiones se añaden en el sumador de tensiones (Σ).



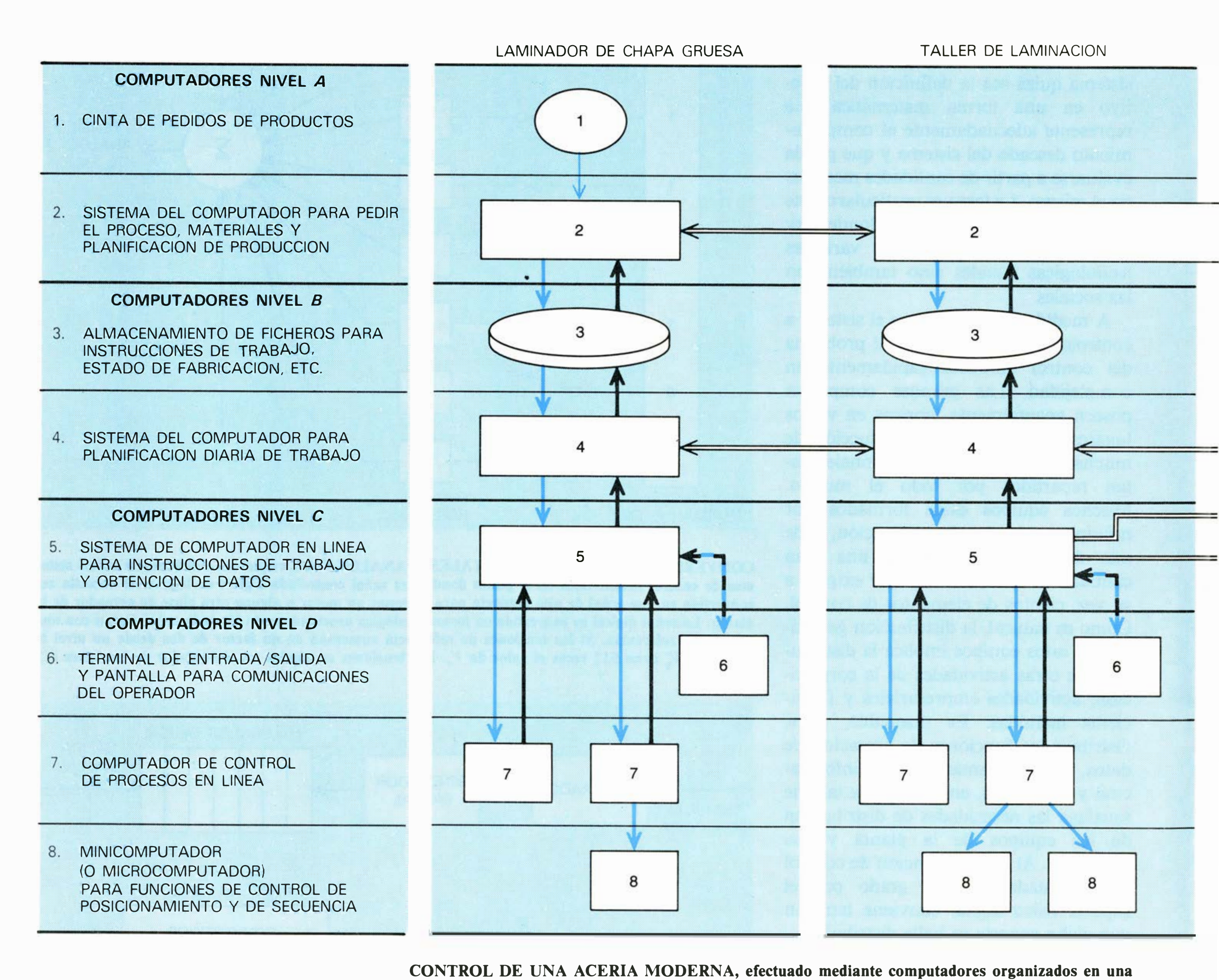
EL CONVERTIDOR ANALOGICO A DIGITAL incorpora en un diseño un lazo de realimentación de modo tal que la señal de salida digital pueda compararse con la de entrada analógica. Si la señal convertida es muy pequeña, el comparador de la salida dispara el generador digital para aumentar el registro digital. En este esquema simplificado, el lazo incorpora un convertidor digital a analógico.

compensar los cambios imprevistos en el comportamiento del sistema más allá de los causados por perturbaciones locales.

Los cambios en el rendimiento total debidos al comportamiento de cada elemento del sistema y a la interacción entre elementos constituyen una dificultad importante para el control de los sistemas complejos. Las acciones de control impuestas en una zona de la planta influyen a menudo en otras. Para integrar el funcionamiento de todos los componentes en una unidad que persiga el objetivo final de la planta, es necesario coordinar las funciones del control distribuido mediante una jerarquización de los computadores de control organizados en sistemas de múltiples niveles.

En la fabricación de acero se da una de las aplicaciones industriales más avanzadas de los computadores en el tratamiento de la información y control en tiempo real. De ahí que recurramos a esa industria metalúrgica en muchos de nuestros ejemplos. Unido ello a que nuestro grupo de investigación relacionado con el control de sistemas complejos, en el Case Institute of Technology, ha realizado cierto número de estudios sobre las necesidades especiales de control de las acerías. La fabricación del acero tiene cuatro fases básicas: la reducción del mineral de hierro para producir hierro fundido, la purificación del hierro para formar acero, la laminación del acero en formas variadas y el tratamiento metalúrgico de éste para satisfacer las especificaciones del producto.

Los requerimientos de control de la laminación del acero en un tren de chapa en caliente son particularmente exigentes debido a la alta velocidad del proceso y a las estrechas tolerancias en las condiciones de operación. El tren de laminación produce la chapa de acero necesaria para la plancha de los automóviles, ciertas aplicaciones domésticas y un gran número de otros productos. El acero entra en la laminación en forma de bloques calentados al rojo de unos 30 centímetros de grosor. Los bloques pasan a través de una serie sucesiva de cajas o conjuntos de rodillos, que van prensando progresivamente el acero en bloques cada vez más delgados. A medida



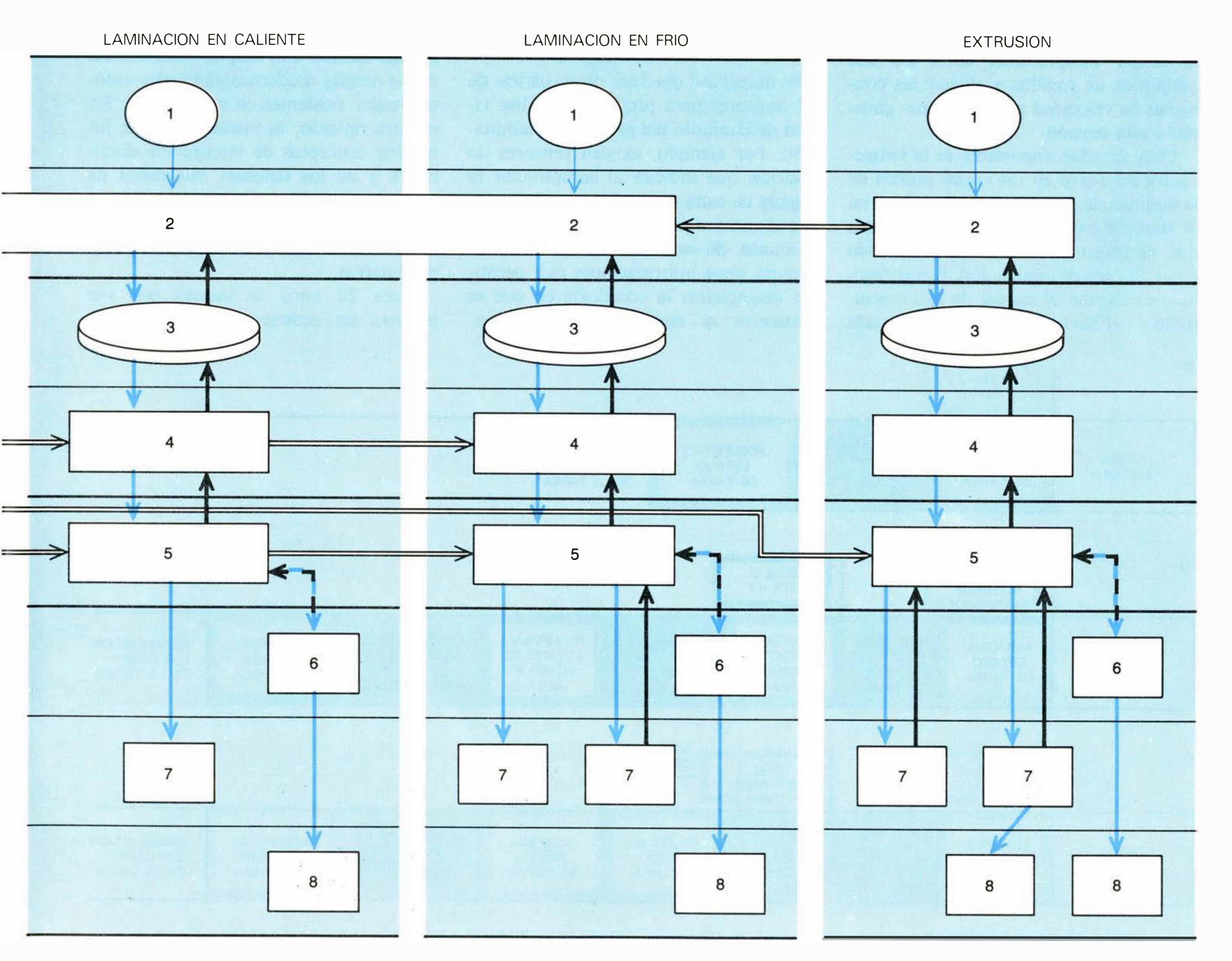
CONTROL DE UNA ACERIA MODERNA, efectuado mediante computadores organizados en una estructura jerárquica. La ilustración es una versión muy simplificada del sistema de control existente en la Kimitsu Works de la Nippon Steel Company, que es la primera realización de una acería altamente computerizada proyectada para el control integrado del sistema total. El control jerárquico de cuarto nivel del tren de laminación comprende la planificación de la producción, la programación, el control de operaciones y el control de procesos, conjuntamente con tareas complementarias tales como las comunicaciones del operador, el mantenimiento del registro, la obtención de datos, el control de inventario y la contabilidad general de la planta industrial. No se muestran los computadores para el control de la

que el acero abandona el último conjunto de rodillos reducido ya a un espesor de un cuarto de milímetro, o menos, su velocidad puede alcanzar los 55 kilómetros por hora.

Un tren de laminación en caliente típico puede disponer de doce o más cajas. El espesor del acero que sale de cada caja está controlado por la separación entre los rodillos. Esta separación está regulada por un controlador que posiciona el rodillo superior por medio de unos potentes tornillos motorizados. Las fuerzas que se desarrollan al comprimir el acero son tan grandes que dilatan el armazón de soporte y doblan los propios rodillos. Debido a ello, el espesor de la chapa de acero que abandona en su recorrido un conjunto de rodillos se apartará significativamente del valor ajustado en la separación de los mismos. Para el control preciso del espesor de la chapa que sale de la última caja del tren de laminación en caliente, es necesario retroalimentar la medida del espesor (proporcionada, por ejemplo, por un medidor de espesores de rayos X) de suerte que puedan ajustarse las posiciones del rodillo para satisfacer las especificaciones de espesor del producto.

En la práctica, no es necesario proporcionar esta clase de realimentación para cada caja de laminación. En cualquier parte de la línea puede estimarse el espesor de la banda de acero con la suficiente precisión a partir de la separación entre rodillos, la presión entre los mismos y otras mediciones de funcionamiento. Lo importante es que la forma de distribución de las reducciones en el espesor de la chapa entre los conjuntos de rodillos influye en la calidad del producto, en la cantidad total de energía requerida y en el ritmo de producción. La distribución de estas reducciones, necesaria para el buen rendimiento de la laminación, viene determinada por un computador situado a un alto nivel en la jerarquía, que toma en consideración el tipo y grado del acero que se está laminando, el espesor final de la chapa, las condiciones de los rodillos y otros múltiples factores.

Además de controlar la separación para cada conjunto de rodillos, hay que regular estrechamente las velocidades de



preparación del mineral, de los hornos de soplado, de la producción de energía y otras actividades. Cada uno de los computadores de control de procesos en el nivel *D* de la figura realiza una variedad de tareas comparable a las ejecutadas por el computador de control en línea en el tren de laminación de chapa en caliente. Este computador controla, entre otras cosas, la carga, la extracción y la temperatura del horno de recalentamiento de la plancha, además de hacer lo propio con la anchura, el espesor y la temperatura de la banda de acero conforme ésta va recorriendo los rodillos del tren de laminación. Los computadores de nivel *D* están situados en el suelo del tren de laminación. Los computadores para los niveles más altos están en el centro de cálculo de la zona de trabajo. El sistema de entrada de órdenes, en la oficina principal, distribuye cintas de órdenes para cada grupo de productos a los sistemas de procesamiento de órdenes en el nivel *A*. Tal división del trabajo "intelectual" entre las múltiples unidades de procesamiento en cooperación se reconoce por los técnicos en informática como control por inteligencia distribuida, tema central del presente trabajo.

rotación de éstos. Una falta de acoplamiento de velocidades entre dos conjuntos sucesivos de rodillos provocaría que la chapa de acero, comprendida entre ellos, o bien se doblase o bien se desgarrase. Cualquiera de estas incidencias daría lugar a un paro costoso de la planta de laminación. El problema se complica por el hecho de que, al prensarse la lámina y volverse más delgada, aumenta proporcionalmente en longitud. Como resultado, cada conjunto de rodillos debe girar a más velocidad que el precedente. Por tanto, la sincronización de las velocidades de los rodillos implica una interacción con los ajustes de su separación. Una forma común de ajustar las consignas de los controladores sucesivos de velocidad es medir la tensión (o el grado de flojedad) que tiene la chapa comprendida entre los dos conjuntos de rodillos y ajustar las consignas de velocidad para mantener constante esta tensión.

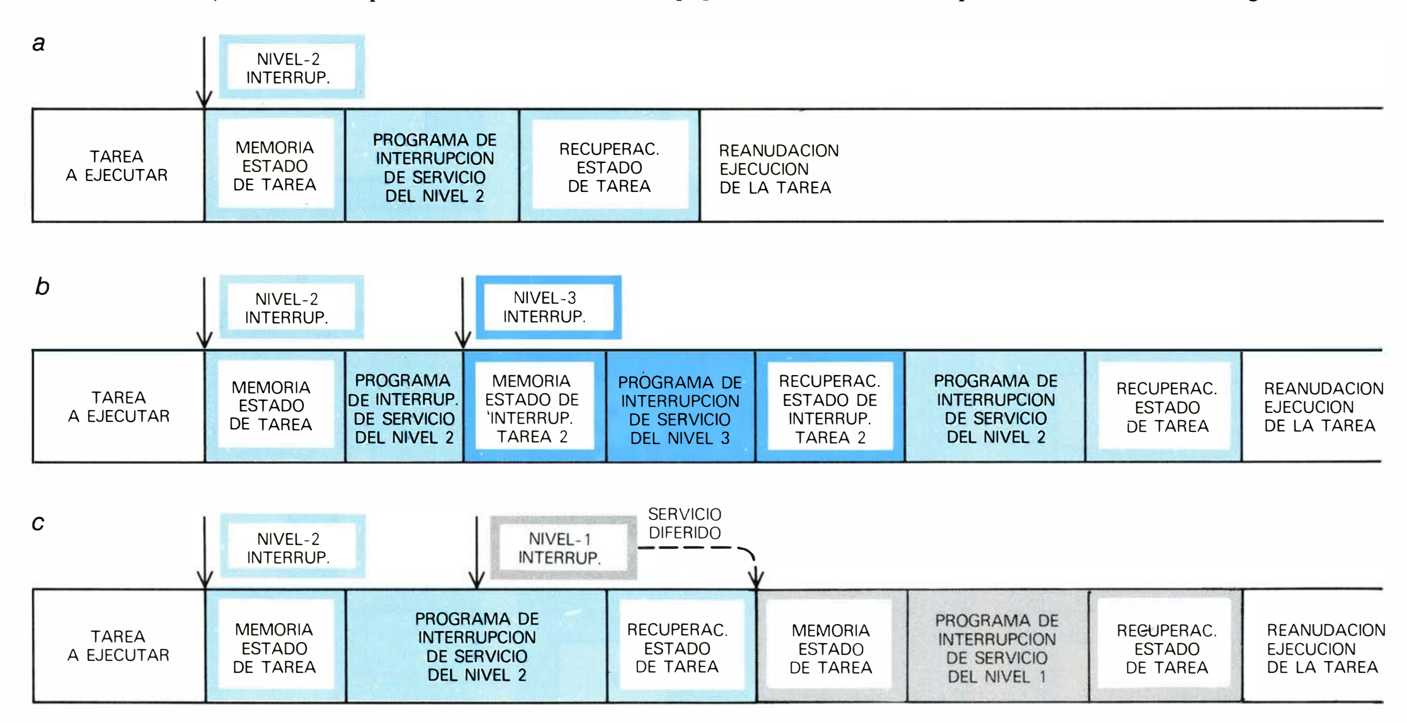
Otra variable importante es la temperatura del acero en los varios puntos de la laminación. Los bloques que entran en el tren deben calentarse primero hasta que alcancen la temperatura requerida en el proceso de laminación. Los controladores ajustan el caudal de fuel consumido en el horno, de modo tal que cada bloque alcanza su temperatura requerida en el tiempo programado para la operación de laminado. Los ciclos de calentamiento pueden prefijarse de tal modo que el consumo de combustible se mantenga al mínimo.

La temperatura de la banda de acero durante la laminación incide en las propiedades metalúrgicas de la chapa, de modo que es necesario vigilar la temperatura de la banda en muchos puntos del tren de laminación. En particular, es muy crítica la temperatura en la zona que se encuentra justo antes de que la banda se arrolle al final del proceso. Por tanto, el control de realimentación se realiza variando el chorro de refrigerante pulverizado aplicado a la chapa para mantener así su temperatura dentro de la zona permisible.

Se necesitan muchas otras clases de sensores para proporcionar una visión de conjunto del proceso de laminación. Por ejemplo, existen sensores de posición que indican al computador la llegada de cada nuevo bloque, de suerte que el sistema pueda iniciar la secuencia adecuada de operaciones. Se obtienen además otras informaciones que permiten diagnosticar la condición en que se encuentra el equipo de laminación. Cuando algo va mal, precisamos información que señale el origen del problema y que especifique qué acciones de emergencia pueden aplicarse hasta tanto no se efectúen las reparaciones correspondientes. Como mínimo, los sensores proporcionan los datos necesarios para evaluar el rendimiento del proceso de laminación, para tener actualizados los coeficientes de los varios algoritmos de control (reglas de procedimiento) empleados y para registrar la historia de la operación desde el punto de vista de contabilización, y de usos analíticos y legales.

Todo lo expuesto anteriormente es interesante porque da una información amplia de la situación crítica en que se encuentra la industria del acero en los Estados Unidos, situación a la que se ha llegado debido a la amplia competencia de las acerías modernas altamente automatizadas existentes en otros países. En nuestra opinión, la introducción de los nuevos conceptos de inteligencia distribuida y de los sistemas integrados de control representa una promesa sustancial para mejorar la situación mediante la instalación de acerías más eficientes y productivas.

Hace 20 años se instaló por vez primera un ordenador digital en una

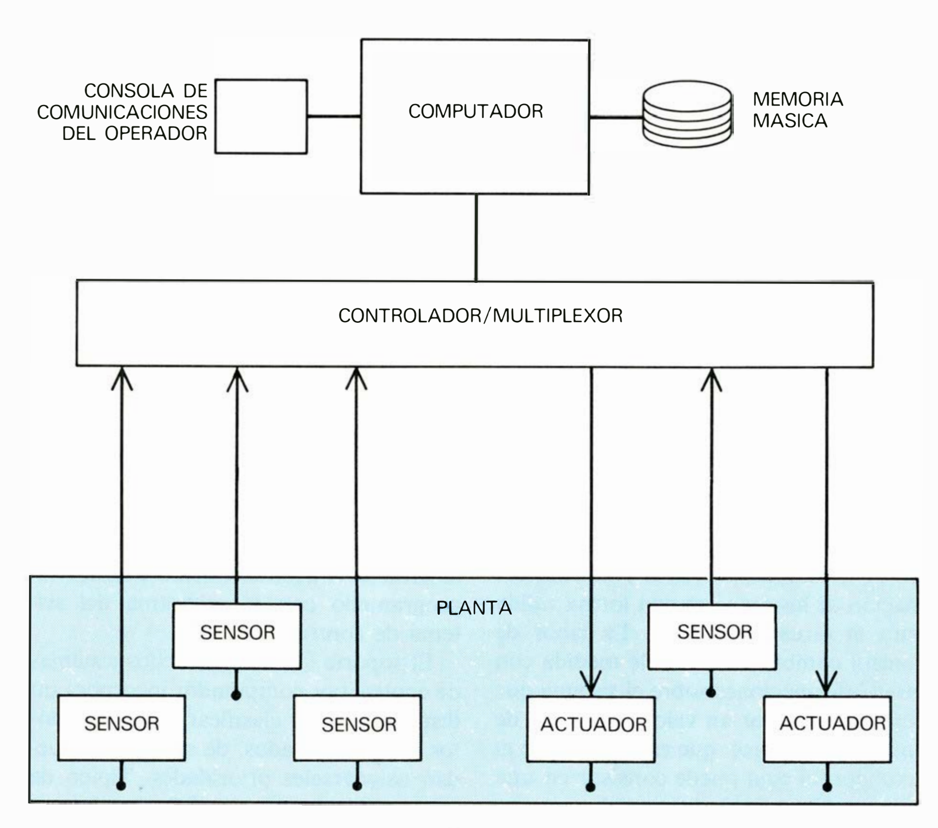


LAS RUTINAS INTERRUMPIDAS son una parte necesaria de los programas en los sistemas de control computerizados. Tales rutinas permiten que un ordenador permanezca en reserva, aparte de su tarea normal para responder a un suceso no programado, como un mal funcionamiento en una pieza del equipo o un fallo de potencia, por ejemplo. A tales emergencias se les asignan prioridades, de forma que un suceso más urgente permitirá interrumpir la atención de otro menos apremiante. En el esquema que se detalla en esta ilustración, el programa responde a una prioridad 2 o nivel 2, e interrumpe el suceso (*a*) almacenando el "estado" (la información contenida) de la tarea normal que se está ejecutando. Después de ser atendido el programa de interrupción del nivel 2, se restablece el estado de la tarea normal y se reanuda su ejecución. La atención de un interruptor de nivel 2 puede interrumpirse por sí mismo a través de un suceso de nivel 3 de alta prioridad. En aquel caso (*b*) el estado de la tarea de interrupción en el nivel 2 es almacenado, mientras que se atiende la interrupción del nivel 3. A continuación se reanuda el mantenimiento del interruptor del nivel 2 seguido por la continuación de la ejecución de la tarea normal. Si en el curso del servicio interrumpido del nivel 2 se presenta una prioridad de un suceso más bajo o de nivel 1, no se tomará ninguna acción hasta que se complete la función de la interrupción del nivel 2 (*c*). (Ilustraciones de Gabor Kiss.)

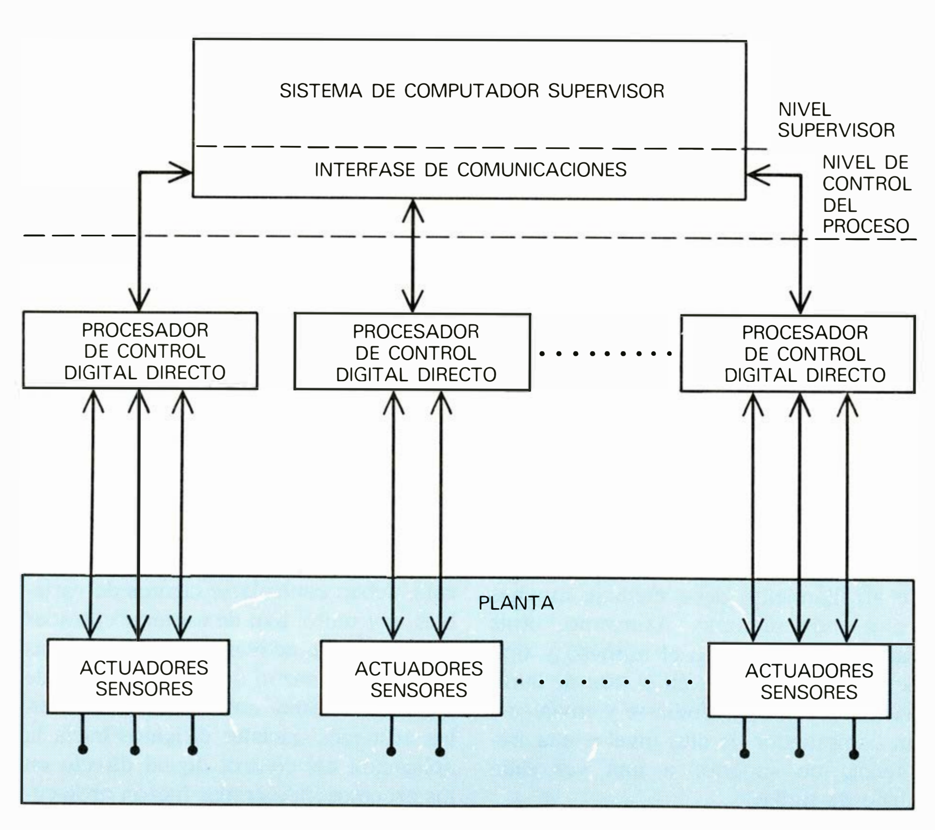
planta industrial. Tenía asignada la tarea de determinar los puntos de consigna (o valores deseados) de los controles analógicos clave, al objeto de mejorar el rendimiento de la planta. El desarrollo alcanzado mereció un gran interés entre los ingenieros de control de la industria y la universidad. Justo antes de este período se inició el programa de investigación del Case Institute of Technology en los sistemas industriales de control, en el que trabajamos en equipo los tres autores de este artículo.

El siguiente paso fue lo que se dio en llamar control digital directo, sistema en el que se sustituyeron muchos controladores tradicionales de proceso por un único ordenador digital. La velocidad del computador le permitió tratar secuencialmente cada lazo de control de realimentación manteniendo el ciclo completo de exploración dentro de las limitaciones de tiempo del proceso controlado. Asimismo, el soporte lógico o programa (software) disponía de potencial suficiente para realizar algoritmos más complejos. Sin embargo, por varios motivos, entre los que se contaban el alto coste y la baja fiabilidad, el control digital directo no fue ampliamente adoptado.

Desde entonces, la tecnología del computador se ha perfeccionado desarrollándose extraordinariamente. Con un índice algo arbitrario (que resume el precio, la longitud de la palabra en código binario, la velocidad, el repertorio de instrucciones, los elementos de entrada-salida y el programa) es posible mostrar que la relación de rendimiento a precio en ordenadores de tamaño medio y en miniordenadores ha aumentado a una velocidad de más del 50 por ciento anual desde los años 1960. Entre 1966 y 1978 el índice rendimiento/precio de los minicomputadores aumentó a más de dos órdenes de magnitud. Además, la programación del lenguaje orientada al usuario ha hecho más fácil a los ingenieros de planta el programar, comprobar y modificar los algoritmos de control. La fiabilidad ha mejorado enormemente como resultado conjunto del incremento de fiabilidad en los componentes del computador y de la introducción de los diseños de tolerancia al fallo, de los componentes redundantes y de las nuevas rutinas para el diagnóstico de averías. Estos avances múltiples en el soporte físico (hardware) y en el programa (software) del computador han aumentado grandemente el dominio de aplicación en aquellos puntos donde es tecnológica y económicamente factible la apli-



ADQUISICION Y CONTROL CENTRALIZADO DE DATOS. La adquisición y el control centralizado de datos pareció constituir un enfoque prometedor en los años 60 al introducirse los grandes computadores de control principal de las compañías IBM, General Electric, Bunker Ramo y otras importantes. Un computador central recibe datos de todos los sensores de la planta, juntamente con las instrucciones del operador y otros datos previamente almacenados en la memoria másica. El computador procesa los datos y remite órdenes a los actuadores. El sistema multiplexor-controlador dispone los datos para introducirlos en forma de embudo en el computador y para la dispersión de las órdenes.



PRIMEROS SISTEMAS de control jerárquico, que se hicieron viables con la aparición del minicomputador hacia finales de los años 60. El control digital directo de las operaciones de la planta podía dividirse entre los minicomputadores; el control supervisor se realizaba en una máquina central.

cación de los computadores a los sistemas industriales de control.

Examinemos ahora con más detalle las tareas básicas que debe realizar un sistema de control por computador y algunos de los problemas del soporte lógico que deben resolverse en el control de un único proceso de la planta. Un sistema de control por computador realiza tres tareas primarias: adquisición de datos, control y actuación. El valor de una variable del proceso (la temperatura, la presión o el caudal, por ejemplo) es transformado por un transductor en una señal eléctrica. La labor de adquisición de datos utilizada en el programa es la conversión de la señal en una forma útil, la comprobación de las cantidades resultantes para asegurar que están dentro de los límites previstos y la transformación de los datos en una forma válida para la tarea de control. La labor de control combina la variable medida con otras informaciones sobre el sistema que permiten calcular un valor de la señal de entrada de control, que es transmitido al actuador (el cual puede consistir en una válvula de control de caudal o en un mecanismo a motor que regule la posición de alguna pieza del sistema).

Puesto que las mediciones de las variables de proceso están usualmente en una forma analógica, y siendo el computador digital un aparato que puede procesar los datos sólo en una forma digital, los datos analógicos deben convertirse a dígitos binarios, o bits (0 y 1). La salida del computador debe reconvertirve entonces de valores digitales a analógicos, para que puedan aceptarse por los actuadores. Tales transformaciones se realizan por piezas especiales del soporte físico, a saber: convertidores analógico-digital y digital-analógico.

La frecuencia a la cual el proceso es interrogado por los aparatos de medida y se somete a acciones de control puede variar ampliamente. Por ejemplo, la temperatura de los bloques de acero en la zona de salida del horno de recalentamiento de un tren de laminación en caliente precisa de un muestreo periódico sólo una vez cada pocos minutos. Por otro lado, la tensión de la chapa de acero que se desplaza entre los soportes de arrollamiento debe medirse muchas veces cada segundo. Asimismo, otras variables, tales como el número y tipo de bloques a prensar en el tren de laminación deben compendiarse y enviarse a un computador de alto nivel a una frecuencia no superior a una vez cada turno de trabajo.

Los sucesos imprevistos que cambian bruscamente el estado del sistema exigen una respuesta inmediata y específica de un determinado computador del sistema. Algunos acontecimientos típicos son el mal funcionamiento de una pieza del equipo, la necesidad de cambiar un conjunto de rodillos en el tren de laminación y la recepción de una orden urgente para programar en el sistema de fabricación. El sistema de control por computador está preparado para manejar tales sucesos por medio de equipos de interrupción. El programa ejecutado corrientemente es interrumpido guardándose la información contenida en aquel momento, usualmente por medio del sistema operativo. Se activa el programa responsable del mantenimiento del estímulo externo; cuando está completo, se reanuda la ejecución del programa interrumpido. De este modo, se sincroniza el acontecimiento exterior no programado con el programa del sistema de control.

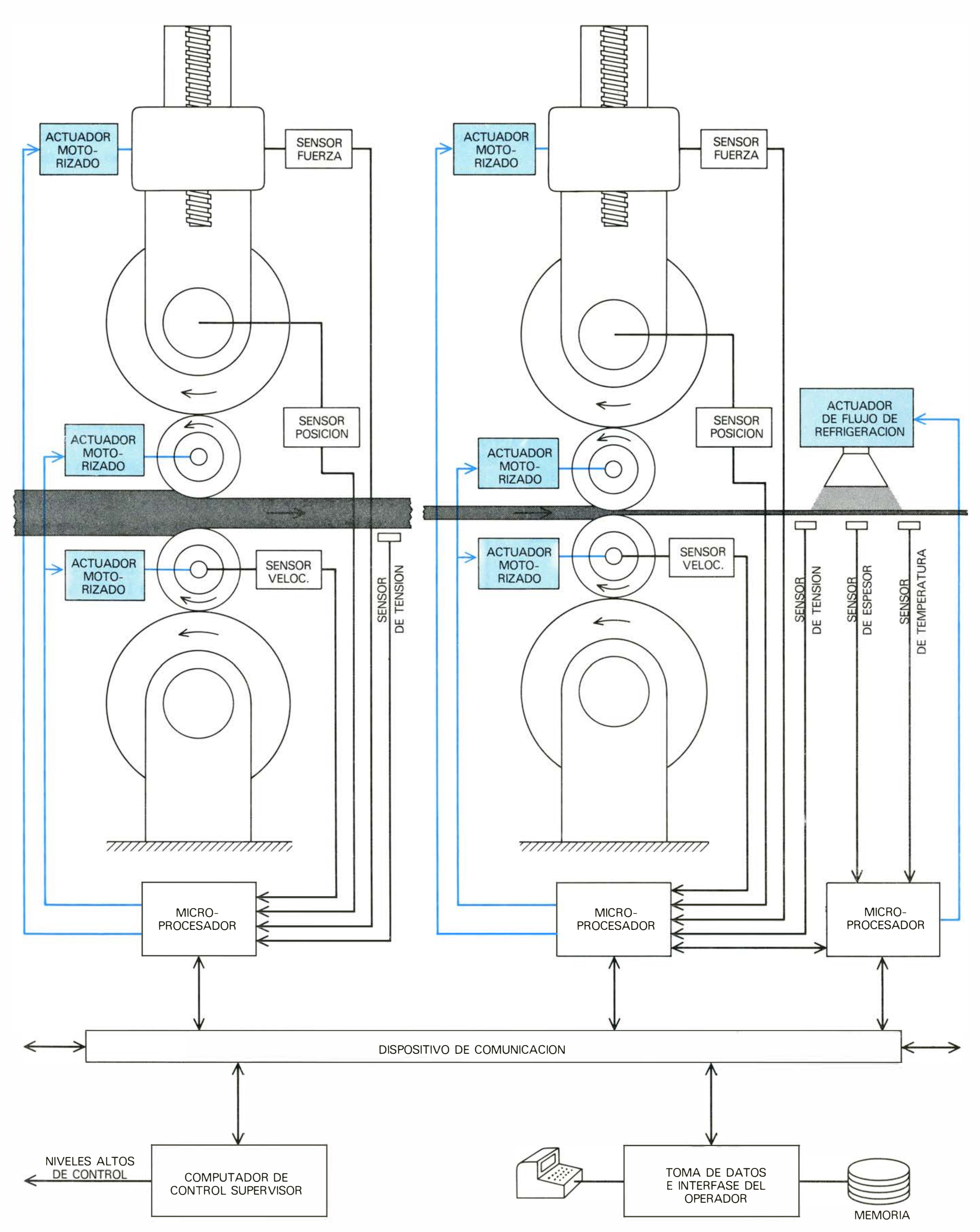
El soporte lógico de muchos sistemas de control por computador incorpora un dispositivo para clasificar acontecimientos no programados, de suerte que puedan asignárseles prioridades. Típico de estas categorías son los fallos en las comunicaciones del operador, fallos de alimentación, alarmas de que ciertas condiciones del proceso presentan un grave error y rutinas para coordinar los aparatos periféricos del computador. En tales categorías se asignan prioridades a los acontecimientos de acuerdo con su urgencia y se les da acceso al sistema de control a través de un sistema de interrupción de niveles múltiples. Por consiguiente, las tareas de interrupción pueden ser interrumpidas, a su vez, por acontecimientos de alta prioridad.

El proceso de marcaje del tiempo en un sistema de control por computador de tiempo real se efectúa conectando el computador a un "reloj" o contador, que interrumpe el computador a intervalos específicos, normalmente cada pocos milisegundos. Manteniendo una suma constante de los sucesos del reloj en la memoria interna del ordenador, el programa de control puede determinar intervalos arbitrarios y tiempos iguales del día.

En procesos industriales a gran escala, deben controlarse cientos de variables; por tanto, han de captarse entradas y tienen que activarse muchas salidas distintas. En razón de la complejidad de las interacciones entre estas variables, los esfuerzos iniciales dirigidos hacia la aplicación del control digital directo en los procesos industriales fueron obstaculizados por la limitada velocidad y por la escasa capacidad de almacenamiento de los sistemas existentes de control por computador y por la falta del soporte lógico adecuado. El primer sistema de control digital directo (realizado hacia el año 1960 en la planta que la compañía química Monsanto tiene instalada en la ciudad de Texas) podía manejar sólo 10 lazos de control debido a las limitaciones del soporte físico, y, en aquel tiempo, los fabricantes de computadores no habían desarrollado virtualmente programas de soporte software. De aquí que se necesitaran varios sistemas de computadores completamente independientes para manejar el escaso número de lazos de control, aún incluso en una planta claramente sencilla. El rendimiento de los sistemas de control digital directo de las plantas iniciales defraudó, debido en gran parte a la ausencia de un soporte lógico eficaz para coordinar los esfuerzos de control de los distintos ordenadores empleados. Estas instalaciones resultaron costosas, no sólo porque se necesitaron varios computadores, sino también porque los directores de la planta insistieron en la necesidad de disponer de controladores analógicos tradicionales para utilizarlos como sistemas de apoyo.

En aquel tiempo, muchos pensaron que la solución final de los problemas que se encontraban en el control digital directo la traería un ordenador central de gran tamaño con la suficiente velocidad, capacidad de memoria y potencia de cálculo para poder manejar todas las variables del proceso. Se argüía también que la centralización del control optimizaría el rendimiento de la planta, gracias a que todos los datos disponibles de ésta se concentrarían en un ordenador único. La introducción de una generación de computadores de control principal de gran tamaño (tales como el IBM 1800, el General Electric 4060, el Bunker Ramo 430 y muchos otros) proporcionó una oportunidad para ensayar el concepto de control centralizado. Como las nuevas máquinas costaban cientos de miles de dólares, sólo valían para grandes instalaciones.

Aunque los grandes computadores demostraron que el control integrado era factible y que la planta mejoraba su rendimiento, introdujeron al mismo tiempo nuevos e intrincados problemas. Uno de ellos, el de la fiabilidad, era muy importante, ya que la mayor parte de la planta era controlada por una única máquina y, en muchos casos, se consideró prohibitivamente cara la instalación de un computador de reserva. (Sin embargo, los computadores redundantes se instalaron en algunas salas de control de unas pocas instalaciones eléctricas de gran capacidad.) Además, el diagnóstico

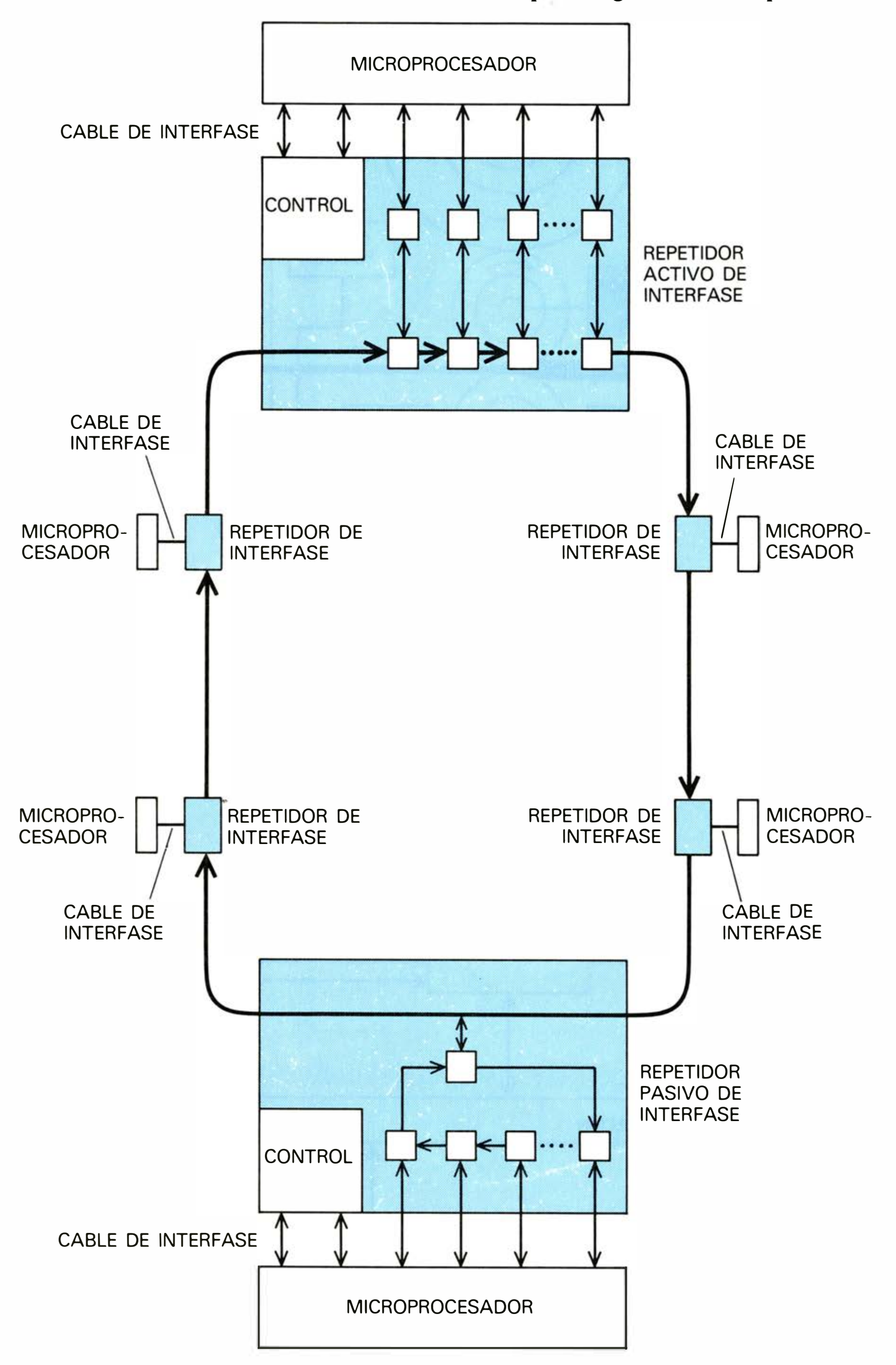


UNA POSIBLE APLICACION de la inteligencia distribuida que se hizo viable por el desarrollo de los microprocesadores, representada para dos cajas de un tren de laminación de acero en caliente. Cada caja del tren de laminación está mantenida por su propio microprocesador, que controla la velocidad y posición de los rodillos en la caja, tomando como base la información sobre las posiciones y velocidades deseadas y verdaderas. Los valores reales son generados por sensores de velocidad y posición; los puntos de consigna (valores deseados) vienen determinados por las medidas de la tensión de la banda y la fuerza de los rodillos y por las entradas recibidas del computador del control supervisor basadas en las especificaciones de producción y en el grado de fabricación del laminador. Puede utilizarse un microprocesador separado (*derecha*) para controlar el caudal refrigerante pulverizado y la separación entre rodillos en el tramo final, a fin de mantener dentro de los valores especificados la temperatura y el espesor de la banda que sale del laminador. Un tren de laminación típico requerirá muchos microprocesadores para el control local, cada uno de los cuales es capaz de comunicarse con otras unidades al mismo nivel de control y con los otros computadores que están operando, a su vez, a niveles más altos.

del software que se precisaba para identificar fallos incipientes de los componentes era virtualmente inexistente. Como resultado de ello, el computador central venía secundado usualmente por sistemas analógicos e incluso manuales. Otro problema con el control centralizado fue la complejidad de los sistemas de entrada y de salida de datos que exigía el multiplexado (interrupción secuencial) de cientos de señales analógicas a través de un canal común de entrada de ordenador. El problema no se resuelve fácilmente, ni siquiera con la electrónica de estado sólido actual; la solución era excesivamente engorrosa y carente de fiabilidad con las técnicas de relé del año 1960.

Un problema más importante, cuya relevancia se ha puesto de manifiesto con el paso del tiempo, es la complejidad del soporte lógico necesaria para controlar grandes sistemas de una forma centralizada. Es necesario desarrollar elaborados programas ejecutivos para dirigir los recursos del control por ordenador e incluirlos entre las tareas competitivas de control en tiempo real. Tales programas ejecutivos en tiempo real tienen cinco funciones principales: programar la unidad central de proceso del computador entre las principales tareas de control, dirigir la memoria principal del computador dividiéndolo entre las tareas a medida que se programan, manejar las operaciones de entrada y salida para evitar conflictos entre las tareas competitivas, mantener la base de datos del sistema (formada por el programa y los ficheros de datos almacenados en la memoria) y proporcionar mecanismos mediante los cuales puedan comunicarse dos o más tareas

Naturalmente, el tamaño y la complejidad de los programas ejecutivos aumentan con el tamaño y la complejidad del sistema a controlar. Los programas ejecutivos para los sistemas de tareas múltiples generalizados hoy en día necesitan por lo general de 18.000 a 25.000 palabras (consistentes típicamente en 16, 24 o 32 bits cada una) de la memoria principal y justifican del 30 al 40 por ciento de todas las instrucciones ejecutadas por el computador. Este gran "techo" influye drásticamente en el rendimiento de todo el sistema. De modo típico, el soporte físico puede responder a una señal interrumpida entre cuatro y diez microsegundos, pero el programa ejecutivo puede aumentar el tiempo de respuesta entre 200 y 300 microsegundos. Por tanto, el tiempo transcurrido entre la captación de un estímulo externo y el momento en que la tarea de control es capaz de responder es de 20 a 80 veces mayor que el tiempo de respuesta básica del computador. Tal retraso es inaceptable en algunas operaciones, de modo que algunos programas ejecutivos incorporan rutinas especiales de exploración que "amortizan" el tiempo máximo necesario para atender una interrupción a lo largo de un gran número de operaciones de entrada y de salida.

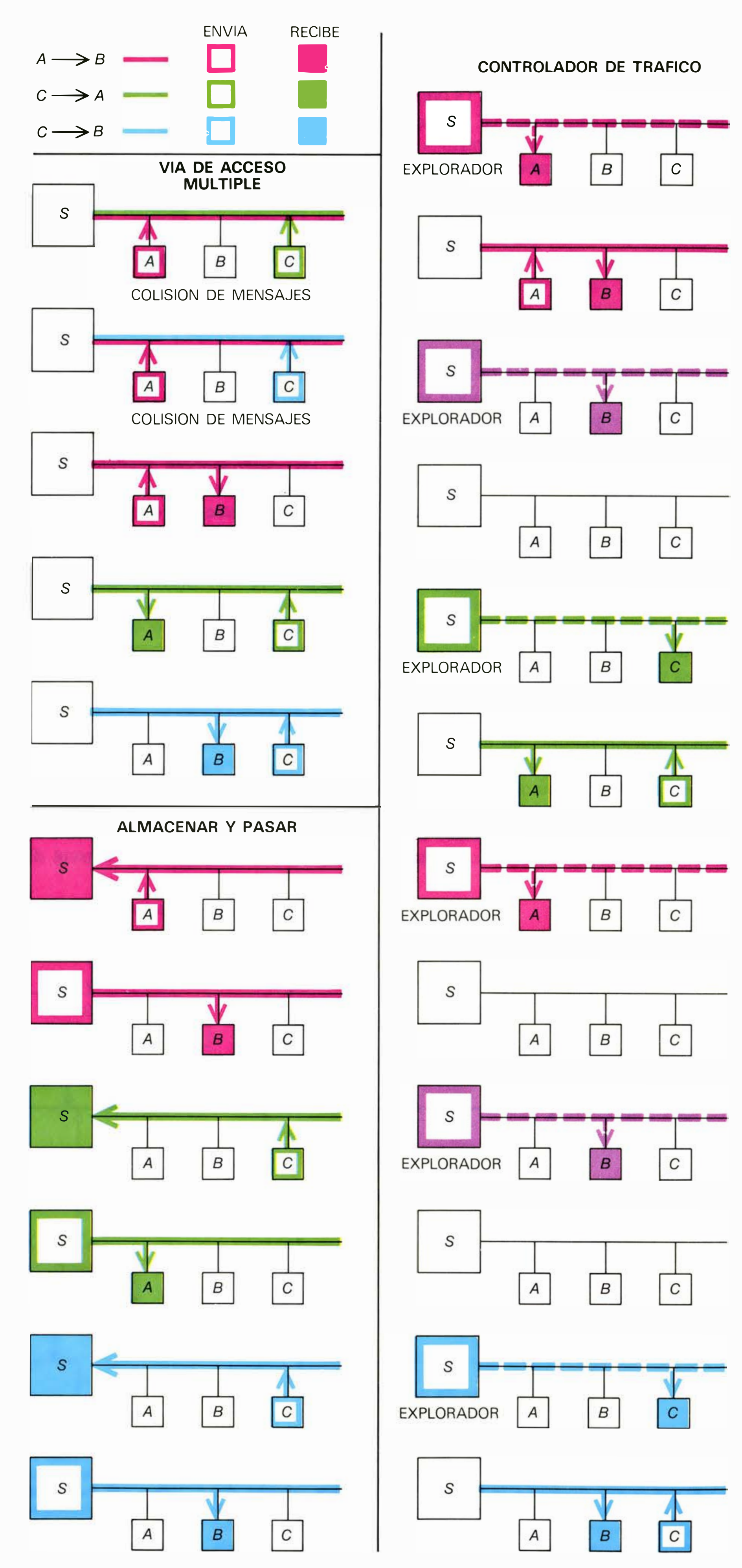


SISTEMA DE COMUNICACION EN FORMA DE ANILLO, de amplio uso para la transmisión de datos a alta velocidad entre microprocesadores. Los elementos de procesamiento en el anillo tienen acceso al canal de comunicación a través de interfases de repetidor, de las cuales existen dos tipos. En realidad, las interfases activas forman parte del anillo. Los bits de datos son cambiados secuencialmente a través de un conjunto de elementos de almacenamiento en la interfase a medida que circulan alrededor del anillo. La red de circuitos de control puede leer mensajes que pasan a su través y puede añadir nuevos mensajes breves insertando más elementos de almacenamiento en la vía de comunicación y alargando temporalmente el anillo. Los repetidores pasivos, a falta de elementos de almacenamiento, se limitan a dar cuenta de los mensajes que los atraviesan. En un esquema se insertan mensajes a intervalos fijos. En otro es posible "radiar" un mensaje hasta su transmisión sin "chocar" con otro.

En los primeros años del decenio 1960-69 varios grupos de investigadores desarrollaron los conceptos de control de nivel múltiple y jerárquico de los procesos industriales. Entre ellos se encontraban Mihajlo Mesarovic, Donald Macko, James D. Schoeffler, Leon Lasdon, Coleman Brosilow y uno de nosotros (Lefkowitz), en el Case Institute, y Theodore J. Williams y su grupo, en la compañía Monsanto y, más tarde, en la

Universidad Purdue. Pero no fue hasta el advenimiento de los miniordenadores de muy bajo precio, hacia el final de la década de 1960, cuando llegó a ser factible la consideración seria de la utilización de un sistema de control por computador jerarquizado. En tal sistema, la supervisión de las operaciones de la planta se repartía entre un cierto número de miniordenadores que rendían cuentas y recibían órdenes desde un gran computador supervisor. Los precios en disminución del soporte físico del ordenador pusieron de manifiesto el coste verdadero para desarrollar el soporte lógico necesario, que en muchos casos se encontró que era de 10 a 20 veces más caro que el soporte físico.

A continuación, en 1970 se introdujo el primer microprocesador, el Intel 4004. Fabricado en una pastilla de silicio de 6 milímetros de lado, el 4004 era un procesador de cuatro bits: una pequeña unidad de procesamiento central por computador capaz de efectuar operaciones aritméticas con cuatro bits al mismo tiempo. Se proyectó originalmente para calculadoras manuales. Al 4004 le siguió el modelo 9900, de 16 bits, de la Texas Instruments, y los modelos todavía más potentes ofrecidos actualmente por media docena de fabricantes. En sólo siete años, entre 1970 y 1977, aparecieron en el mercado tres generaciones completas de microprocesadores. Los modelos corrientes son auténticos microordenadores, con una unidad de procesamiento central, una memoria y dispositivos de entrada y salida en la misma pastilla. En los últimos siete meses, tres fabricantes han anunciado la



VIA GLOBAL COMPARTIDA con distintas variaciones de acceso; se emplea también ampliamente como medio de comunicación para enlazar microprocesadores. La vía proporciona un camino de dos direcciones a las series de datos. En la figura están representados tres esquemas de acceso. En el ejemplo que se muestra aquí la secuencia deseada de comunicaciones desde un microprocesador a otro es *A* a *B*, *C* a *A* y *C* a *B*. En el esquema más sencillo, conocido como sensor transportador de acceso múltiple, las interfases pasivas que tienen mensajes para enviar comprueban la vía para detectar la presencia de un mensaje en camino. Si no existe ningún otro mensaje, radian el suyo ya sea inmediatamente o después de un intervalo al azar. Si dos o más interfases transmiten al mismo tiempo, se presenta una colisión y se pierden ambos mensajes. A continuación, cada interfase espera un intervalo libre y retransmite. En algunos diseños, se incluye un "árbitro de acceso" para determinar qué interfase puede utilizar la vía de acuerdo con alguna regla prioritaria. En un segundo esquema, puede utilizarse un interruptor para almacenar y enviar mensajes de un aparato al otro ($A \rightarrow S$, $S \rightarrow B$, y así sucesivamente). En un tercer esquema, un "controlador de tráfico" vigila cada interfase para saber si tiene mensajes que transmitir. Si el controlador escruta *A*($S \rightarrow A$) y halla que *A* tiene un mensaje, a *A* se le da un acceso a la vía ($A \rightarrow B$).

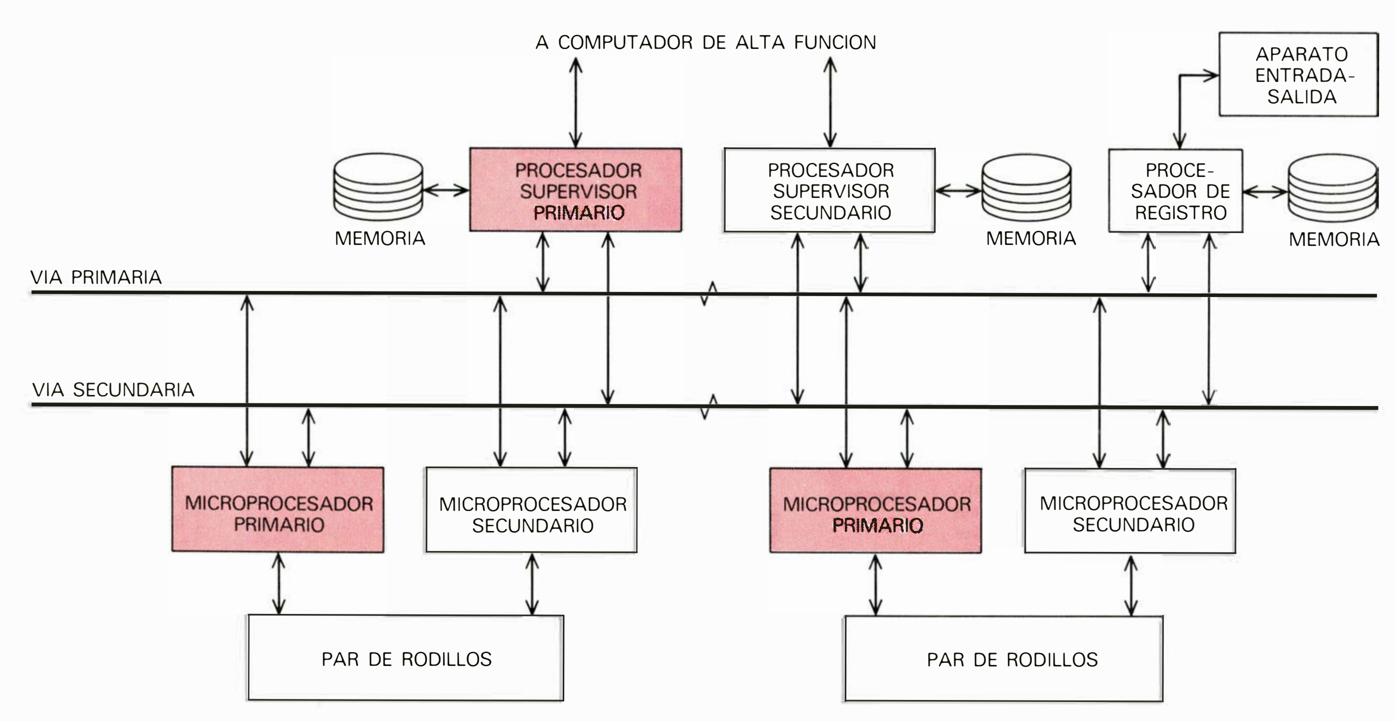
aparición de microprocesadores con un número de puertas entre 15.000 y 20.000 en una simple pastilla. (Se llama puerta a un conjunto de elementos del circuito que llevan a cabo una operación lógica.) La potencia de procesamiento de las nuevas pastillas supera la de muchos miniordenadores de alto rendimiento. Dentro de tres años, tales pastillas se venderán por menos de 30 dólares y un computador que contenga una de ellas conjuntamente con una memoria de un millón de palabras de ocho bits costará posiblemente un precio inferior a 20.000 dólares. Hoy en día esta memoria separada se vende por unos 35.000 dólares formando parte de un sistema de miniordenadores.

La nueva tecnología del microprocesador llamó inmediatamente la atención de los proyectistas que empezaron a considerar el impacto que produciría un ordenador de bajo coste y relativamente potente al aplicarlo a resolver los problemas del soporte lógico de las entradas y salidas y de la fiabilidad. Como la potencia absoluta de cálculo de cada microprocesador no es muy grande, se piensa, como solución natural, en una disposición arquitectónica múltiple basada en la inteligencia distribuida. Si tal arquitectura se aplicara al control del tren de laminador de chapa en caliente, cada conjunto de rodillos estaría controlado por su propio microprocesador. Las entradas al procesador podrían dar cuenta de la posición del rodillo y de la presión; las salidas podrían controlar el funcionamiento de los motores que determinan la posición y la velocidad de los rodillos. Otros procesadores podrían encargarse de determinar el espesor, la temperatura y la tensión de la banda y controlar el caudal del chorro de refrigerante pulverizado. Los puntos de consigna de todos los lazos de control pueden calcularse mediante un procesador de supervisión tomando como base los algoritmos diseñados para optimizar el rendimiento del tren de laminación. Como cada procesador manejaría menos lazos de control que los controladores digitales directos de los sistemas antiguos, la multiplexión de las señales de entrada y salida dejaría de ser ya un problema importante.

Se han dedicado muchos esfuerzos en orden a descubrir las clases de medios de comunicación que podrían resultar más eficaces para interconectar los microprocesadores de un sistema de control distribuido. El tamaño de las plantas y el coste del cableado ha obligado a que prácticamente todos los sistemas distribuidos estén situados alrededor de un sistema de comunicaciones de alta velocidad en el cual cada bit de información es transmitido secuencialmente sobre un canal único. En el nivel de control directo, todos los elementos de procesamiento suelen comunicarse a través de un canal de este tipo. El canal puede constar de un simple par de hilos retorcidos o, si se requieren altas velocidades de transmisión, de un cable coaxial o ser un enlace por fibra óptica. Los esquemas alternativos en los cuales se transmiten simultáneamente varios flujos de bits en varios canales cuestan mucho más dinero.

Se han propuesto unos diez esquemas arquitectónicos para los sistemas de inteligencia distribuida, pero sólo se han aplicado ampliamente tres en la adquisición de datos y en el control: el sistema en anillo (o lazo), la vía global compartida y la vía global interrumpida. En el sistema en anillo, los elementos de procesamiento y los aparatos de entrada y salida se conectan al anillo por simples aparatos de interfase que interceptan cualquier mensaje que lleve la dirección adecuada. Los aparatos de interfase pueden ser activos o pasivos. Las interfases activas forman parte del anillo, y de este modo cada bit de cada mensaje pasa a su través. Las interfases pasivas no forman parte del anillo: sólo copian adecuadamente los mensajes dirigidos a medida que éstos circulan.

Los aparatos de interfase deben ser capaces también de insertar mensajes en el anillo. Se encuentran en servicio tres esquemas generales. En uno de ellos,



MICROPROCESADORES Y VIAS REDUNDANTES, usuales en el control digital de procesos industriales. Esta ilustración muestra esquemáticamente una posible disposición de medios duplicados para controlar dos cajas de un tren de laminación de acero en caliente. Se asignan microprocesadores redundantes a cada par de rodillos en una caja. El procesador secundario pregunta periódicamente a la unidad primaria si de hecho ¿está usted bien? Recibe también mesajes duplicados del punto de consigna desde el elemento de procesamiento supervisor y asegura periódicamente al supervisor que está realmente bien. Si un microprocesador primario falla, la unidad secundaria lo notifica al supervisor y asume el control de los rodillos. El supervisor secundario se encarga del control en caso de que falle el primario. Si falla la vía primaria, la secundaria es conmutada en seguida.

cada interfase tiene un intervalo de tiempo asignado durante el cual puede insertar un mensaje. En un segundo esquema se transmiten bits de datos "nulos" (secuencias de ocho ceros) en el anillo cuando en él no existe tráfico de señales. Al detectar un bit nulo, una interfase puede insertar un mensaje. En un tercer esquema, una interfase que tiene un mensaje para el anillo, "radia" el mensaje de forma repetitiva. El sistema está proyectado de tal forma que, dentro de un período garantizado estadísticamente, cada mensaje se transmitirá libre de interferencias de otro mensaje. El tercer esquema requiere que el anillo trabaje a una velocidad de transmisión de bits mucho más alta que la requerida por los dos primeros esquemas, caso de que se desee alcanzar la misma velocidad de transmisión del mensaje porque muchos de ellos serán desechados por solapamiento con los otros mensajes. Sin embargo, el esquema tiene la gran ventaja de que las interfases pueden ser muy sencillas y por tanto fiables.

La vía global compartida es una vía abierta, y no cerrada como era el caso de un anillo. Posee una ventaja importante sobre el anillo al transmitir mensajes en dos direcciones. La interfase se alcanza de la misma forma que en el anillo. La inserción radiada de los mensajes es una práctica común. En algunos sistemas se incluyen líneas adicionales para permitir el arbitraje automático de las demandas para el acceso a la vía. Alternativamente, el aparato de arbitraje puede supervisar el acceso a la vía mediante el escrutinio periódico de los aparatos de interfase para determinar si alguno de ellos tiene mensajes que transmitir.

En la vía interrumpida, los mensajes desde un microprocesador al otro se envían usualmente a un interruptor que los dirige al destino adecuado. El escrutinio es práctica común. Un sistema recientemente desarrollado, con un enfoque del todo distinto, posee como paso de comunicación un cable coaxial de alta capacidad. El cable puede transportar una banda amplia de frecuencias, de modo que cada procesador puede transmitir mensajes al interruptor a una frecuencia única y evitar así el problema del arbitraje. El interruptor almacena los mensajes y los retransmite al procesador de destino a la frecuencia única propia del microprocesador. De esta suerte pueden transmitirse a la vez muchos canales de datos, de televisión y de voz. La técnica se denomina multiplexión por división de frecuencia.

Entre los atributos importantes de cualquier sistema de control de inteligencia distribuida se halla la modularidad, la tolerancia a fallos y la fiabilidad. El término modularidad significa que si el sistema en anillo o en vía tiene una adecuada capacidad de dirección, se puede añadir, casi en cualquier parte y a un coste bajo, sensores adicionales, actuadores y microprocesadores. Los aparatos de interfase tienen una lógica simple y son baratos comparados con los aparatos que interconectan. Por tanto, la modularidad no es difícil de alcanzar.

En cuanto a la tolerancia a fallos y a la fiabilidad, las vías y las interfases de un sistema de inteligencia distribuida están instalados normalmente por duplicado para prever el fallo en las comunicaciones. Luego, en caso de que se dé un fallo, los sistemas de apoyo pueden conmutarse automáticamente. Hoy en día, los microprocesadores son inherentemente seguros, pero también pueden fallar. Un esquema para superar el posible fallo de un procesador es instalar sensores y actuadores por duplicado y enlazarlos con dos procesadores diferentes, de los cuales uno sólo está operando activamente. Si falla el microprocesador de operación, actúa el procesador de apoyo, contando con sus propias conexiones a los sensores y actuadores. Una alternativa es proporcionar un sistema de conmutación, de forma que, en caso de fallo del procesador, los datos del sensor sean conmutados a un procesador alternativo, al cual se le da el acceso simultáneo a los actuadores de la unidad que falla.

Si falla un microprocesador situado a cualquier nivel en la jerarquía por encima del nivel más bajo del procesador, otro procesador (u otros) puede asumir su carga en su propio nivel, siempre que la memoria de la unidad de apoyo contenga una copia de las tareas a asumir. Como los mensajes entre tareas en un sistema jerárquico se envían a través de los medios de comunicación, cada mensaje debe incluir la dirección de la tarea destinada. La dirección puede ser la del procesador que se descarga de la tarea o el nombre (la designación numérica) de la propia tarea. En el primer instante, el fallo de un procesador y la reconfiguración consiguiente del sistema cambiaría las direcciones de algunas tareas. Entonces, todos los cambios habidos deberían notificarse a todas las tareas, ya que los mensajes tendrán que encaminarse a las nuevas direcciones. Esta solución es claramente chapucera.

La alternativa preferible es proporcio-

nar a cada interfase de comunicación todos los nombres de las tareas asignadas al microprocesador al cual está conectada. A medida que los mensajes circulan por la línea, cada interfase reconoce e intercepta aquellos que pertenecen a sus propias tareas. Este esquema manejará también los fallos en los procesadores de adquisición de datos. Mediante el envío periódico mutuo de ¿está usted bien? los procesadores de mensajes pueden determinar cuándo un procesador vecino ha fallado e iniciar las reconfiguraciones que redistribuirán la carga de trabajo. Este enfoque recibe el nombre de dirección suave de las tareas.

La información contenida en un microprocesador en el momento del fallo se conoce como estado del procesador. ¿Cómo puede recuperarse el estado del procesador estropeado de modo que pueda mantenerse un funcionamiento adecuado del sistema? No se ha encontrado una solución general del problema. Sin lugar a dudas, se han ideado soluciones especializadas y verdaderamente han trabajado en instantes específicos.

Las vías de comunicación en un sistema de control de inteligencia distribuida están expuestas a menudo a ambientes hostiles (motores eléctricos, campos magnéticos, etcétera) de modo que, a menudo, los mensajes del sistema se desechan porque se han alterado uno o más bits por un ruido eléctrico o de otro tipo. Existen varios esquemas para añadir bits redundantes a cada mensaje de suerte que el receptor pueda detectar un error a la llegada del mensaje. Si se detecta un error, se pregunta al aparato emisor que repita el mensaje hasta que se reciba una versión limpia. Ciertos esquemas adicionales protegen contra la pérdida o la duplicación de los mensajes. No sólo permiten detectar los errores en el receptor sino incluso corregirlos.

La aportación de la capacidad adecuada a un sistema de control de inteligencia distribuida no suele ser difícil porque el proyectista del sistema generalmente conoce bien y por anticipado las velocidades de tráfico de mensajes y los requerimientos en los tiempos de respuesta. Es posible diseñar los medios de comunicación, el número de microprocesadores y la asignación de las tareas a los procesadores con el grado necesario de exceso de capacidad. El rendimiento total del sistema es influido de forma interesante por la distribución del tratamiento. Limitando las funciones de los procesadores de bajo nivel al control digital directo de un pequeño número de lazos de control, puede reducirse el número de tareas que cada procesador debe ejecutar a un nivel que permite completar cada conjunto de tareas sin ser asegurada su prioridad para tareas más importantes o más urgentes. De aquí que la estructura del sistema de operación pueda simplificarse significativamente. El rendimiento viene también reforzado por el hecho de que las tareas en un sistema distribuido pueden realizarse con preferencia en paralelo, o en conjunto, más que secuencialmente, tal como se efectuarían en un único ordenador de gran tamaño.

Por otro lado, al no estar centralizadas las tareas, los mensajes entre tareas dan lugar a retrasos en la transmisión y en el tratamiento. Además, cuando un procesador de adquisición de datos de bajo nivel detecta un suceso externo "nocivo", puede transcurrir una fracción sustancial de segundo antes de que la alarma pueda pasar hacia adelante para iniciar un programa interrumpido en un procesador de más alto nivel, si éste realmente está donde debe generarse la respuesta. Tales desventajas pueden hacerse mínimas mediante la asignación al mismo procesador de tareas que se interaccionen frecuentemente. Esto se realiza con mayor facilidad dedicando procesadores a adquisición de datos de bajo nivel y a funciones de control directo, de modo que sólo sea necesario transmitir alarmas e información resumida a los procesadores de alto nivel.

El avance explosivo de la tecnología de los semiconductores continuará conduciendo la evolución de los sistemas de control y les permitirá utilizar con ventaja los beneficios de la inteligencia distribuida. Con esta tecnología se puede comenzar a explotar la teoría del control jerárquico, pero existe todavía el hecho de que la teoría actual disponible para el proyecto de tales sistemas de control no ha progresado lo suficiente como para que pueda aplicarse en la consecución del control óptimo de los equipos industriales complejos. Nosotros, en el Case Institute, para remediar esta deficiencia, y contando con la ayuda de nuestros colegas y de los estudiantes, estamos desarrollando metodologías para modelar y proyectar sistemas de inteligencia distribuida aplicables a procesos industriales variados. Un nuevo laboratorio establecido con el soporte de la National Science Foundation y de la compañía Foxboro está desempeñando un papel central en su esfuerzo de desafío, para estudiar el control distribuido por computador.

Ciencia y sociedad

Sexafs

Nuestro conocimiento de gran parte de la estructura atómica de la materia, sobre la que se basa la física, la química y la biología modernas, está íntimamente ligado a los métodos de difracción, tanto de rayos X como de electrones. Un problema específico de este terreno, y de gran actualidad, es la determinación de estructuras atómicas superficiales. Un cabal entendimiento de la posición que ocupan los átomos en las superficies de los sólidos se hace imprescindible para la comprensión de problemas tales como el enlace químico en la superficie, la oxidación y la catálisis, algunos de ellos de notable relevancia tecnológica.

La difracción de electrones de baja energía (que se abrevia con el acrónimo LEED, del inglés "Low Energy Electron Diffraction") es hoy la técnica por antonomasia seguida para la determinación de estructuras superficiales [véase "Estructuras atómicas de superficies cristalinas", de Pedro M. Echenique y Michel A. Van Hove, en *Investigación y Ciencia,* abril 1979, págs. 28-38].

Recientemente, sin embargo, un artículo de la prestigiosa revista inglesa *Nature* [vol. 278, 19 de abril de 1979] llamaba la atención sobre un importante experimento realizado por Bianconi y Bachrach, publicado en *Physical Review Letters* [42, 104, 1979]. Dichos investigadores han demostrado, por primera vez, que la técnica conocida como SEXAFS (del inglés "Surface extended X-Ray Absorption Fine Structure") puede llegar a ser una alternativa poderosa al LEED en la investigación de la geometría superficial de superficies cristalinas. El método SEXAFS consiste en lanzar rayos X sobre el sólido y observar los *electrones* que emergen. La contrapartida del SEXAFS para el estudio del interior de los sólidos, EXAFS, consiste en medir no los electrones que emergen sino el flujo de rayos X que han perdido energía, es decir, el coeficiente de absorción. Esta técnica, EXAFS, es un complemento poderoso para la difracción de rayos X, y, como veremos, proporciona una información complementaria particularmente necesaria en aquellos casos en que no hay orden de largo alcance.

En un material con orden de largo alcance todos los átomos están situados en posiciones tales que su estructura geométrica siempre se repite. No importa cuán lejos se mueva uno en una dirección determinada, seguirá encontrando los átomos según intervalos periódicos fijos. Si esto no ocurre, el material no tiene orden de largo alcance; este es el caso, por ejemplo, de los materiales amorfos. Dicho de otra manera, un átomo suficientemente alejado de un punto determinado en un material amorfo tiene la misma probabilidad de encontrarse a cualquier distancia.

El principio físico clave involucrado en SEXAFS es esencialmente el mismo que en EXAFS. Este último encierra un gran interés y ha adquirido notable importancia recientemente. Para comprender tanto SEXAFS como EXAFS, examinemos primero los mecanismos de interacción de los rayos X con la materia condensada. La radiación X interacciona fundamentalmente con los electrones del medio. La teoría cuántica nos dice que toda la radiación electromagnética está compuesta de partículas llamadas fotones cuya energía es proporcional a la frecuencia de la radiación. La constante de proporcionalidad es la famosa constante de Planck. Por otro lado, la mecánica cuántica también nos dice que los niveles de energía de los electrones están cuantificados, es decir, un electrón en un átomo sólo puede tener unas energías determinadas E_1, E_2, E_3 y no una energía intermedia.

Cuando un haz de rayos X incide sobre un átomo puede ceder su energía a un electrón que, como resultado de la colisión, puede quedar expulsado del átomo. En virtud de este proceso, el flujo de rayos X incidente es atenuado, atenuación que se expresa en términos de una cantidad llamada "coeficiente de absorción", que en general tiene una dependencia muy suave con la energía excepto en los llamados bordes de absorción. [En la ilustración de la página opuesta se ofrece una representación esquemática de parte del espectro de absorción del hierro. El coeficiente de absorción nos dice cuántos fotones han desaparecido del flujo inicial por haber cedido su energía a un electrón del medio. Ocurren crecimientos drásticos, (*a*), cuando un nuevo modo de absorción entra en juego. En la figura aparece el llamado borde *K* de absorción del hierro, correspondiente a aquellas energías a las cuales un rayo X puede empezar a liberar electrones del nivel *K*. En *b* se muestra un aumento de la parte de la figura (*a*) cerca del borde de absorción. Las oscilaciones en la gráfica se deben a la interferencia de la onda emergente del átomo que ha absorbido el rayo X y las originadas en los átomos vecinos por colisión de éstos con la onda incidente. En dicha gráfica está contenida la información sobre el entorno atómico de uno dado.]

Cada vez que un rayo X tiene energía suficiente para excitar electrones de un nuevo nivel profundo se produce un brusco incremento del coeficiente de absorción como corresponde a la existencia de un nuevo proceso de pérdida de energía antes inexistente. Excepto por este crecimiento drástico, la variación del coeficiente de absorción por átomos aislados con la energía es muy suave [*véase la ilustración de la derecha, a*]. El espectro de absorción de un sólido, a diferencia de un átomo aislado, presenta una estructura fina muy compleja por encima del borde de absorción (*b*). Esta estructura fina es originada claramente por los átomos que rodean al que absorbe el rayo X; y ésta es la única diferencia de un átomo en el sólido con respecto a un átomo aislado, al menos en lo que se refiere a los niveles profundos en los que estamos interesados.

Los niveles de energía de un átomo aislado son esencialmente los mismos cuando el átomo se encuentra aislado que cuando se encuentra en el interior del sólido rodeado por átomos a distancias cercanas. (Excepto por los más externos, que son los más débilmente ligados, cuya energía es mucho menor que la de un fotón de frecuencia X y por tanto son irrelevantes como veremos.)

La mecánica cuántica nos introduce en el conocimiento de esa estructura fina. Tal como fue sugerido por De Broglie en 1924 y confirmado experimentalmente por Davisson y Germer en 1927, toda partícula tiene asociada a ella una onda cuya longitud de onda depende inversamente de su momento [véase el artículo citado de *Investigación y Ciencia,* pág. 29]. La probabilidad de que un fotón X sea absorbido depende del estado inicial en que se encuentra el electrón y del estado final en que se halla después de adquirir la energía del fotón absorbido. El estado inicial, tal como hemos indicado anteriormente, es esencialmente el mismo para átomos aislados que para átomos en el interior de un sólido (excepto quizá para los átomos más ligeros como el hidrógeno y el helio). El estado final del electrón en el caso de un

átomo aislado puede ser visualizado como una onda esférica que se propaga hacia el exterior de dicho átomo. Si el átomo está rodeado por otros átomos, como es el caso del sólido, la onda colisionará con dichos átomos originando ondas que emanan de ellos. Este fenómeno es análogo a lo que ocurre cuando las ondulaciones producidas por una piedra que cae en un estanque se encuentran con pelotas de tenis flotando en la superficie.

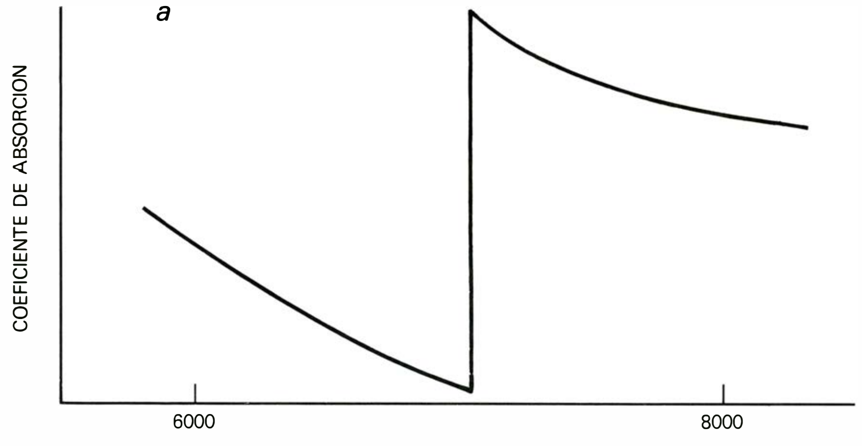
El estado final del electrón es el resultado de la superposición de dichas ondas, que pueden superponerse constructiva o destructivamente según que los máximos de los diferentes frentes de onda coincidan o no. Este fenómeno es el origen de la estructura fina del espectro de absorción de rayos X y depende tanto de la longitud de onda del electrón emergente (y por tanto de la del rayo X incidente) como de la posición de los átomos vecinos con respecto a él, proporcionando por tanto información sobre la estructura atómica de los sólidos. Nos dice cuál es el entorno atómico, regular o no, de un determinado átomo. A pesar de que la existencia de este fenómeno se ha conocido casi durante medio siglo, su potencia como instrumento de análisis estructural tiene su origen en un artículo de Dale E. Sayers, Edward A. Stern y Farrel W. Lytle aparecido en 1971 (D. E. Sayers, E. A. Stern and F. W. Leytte, 1971, *Phys. Rev. Lett., 27,* 1204).

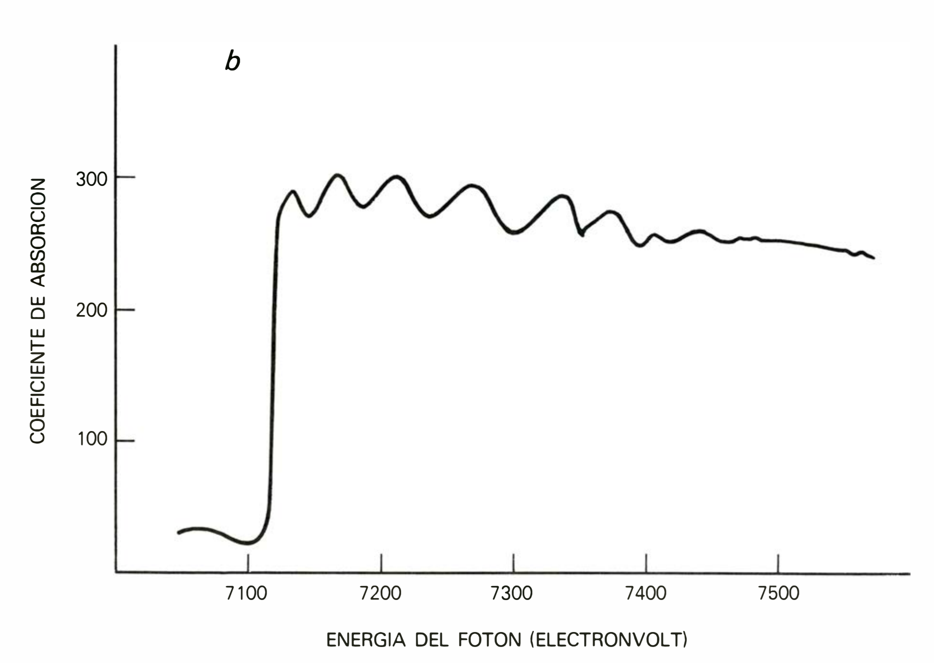
La dificultad de conseguir fuentes adecuadas de rayos X fue la causa de que hasta los últimos años no se pudiesen realizar experimentos de este tipo. La aparición de la radiación de sincrotrón permite realizar estos experimentos de una forma eficiente y rápida, abriendo así un infinito número de posibilidades de investigación, tanto en materiales cristalinos y amorfos como en medios biológicos y superficies.

La radiación de sincrotrón tiene su origen en la radiación que emite una partícula cargada cuando se acelera. Esta propiedad está contenida en las ecuaciones del electromagnetismo formuladas en el siglo XIX por James Clerk Maxwell (precisamente, en este año 1979 celebramos el primer centenario de su muerte). Aceleración es cualquier cambio de velocidad, no sólo de su valor numérico sino de su dirección. Esta es la característica que la radiación de sincrotrón aprovecha. La radiación de sincrotrón es la radiación emitida por un electrón que se mueve con velocidad constante en una trayectoria circular. Sobre él sólo se ejerce una aceleración particularmente sencilla: una aceleración centrípeta constante; por tanto, el electrón radia de una forma continua. La radiación emitida alcanza su máxima intensidad en el plano tangente a la órbita del electrón y tiene una serie de características que le convierten en extremadamente útil para el estudio de superficies, medios sólidos y materiales biológicos. La primera es la *gran intensidad* que se puede obtener a lo largo de un amplio margen del espectro, precisamente el de longitudes de onda comparables a las dimensiones atómicas, radiación muy difícil de obtener con intensidad suficiente por otros medios. Otra ventaja es el alto grado de polarización de la radiación, es decir, el hecho de que todas las ondas electromagnéticas tengan el vector eléctrico asociado a ellas paralelo al plano de la órbita del electrón, lo cual facilita la interpretación de los experimentos y abre la posibilidad de otros nuevos. Aparte de las ventajas citadas, la fuente de radiación de sincrotrón opera en ultra alto vacío; por tanto, es directamente compatible con los requisitos necesarios para efectuar experimentos en superficies no contaminadas.

En un artículo clave, aparecido recientemente en *Phys. Rev. Lett.* 42, 104 (1979) Brianconi y Bachrach han demostrado por primera vez que pueden utilizarse medidas del espectro fino de absorción de rayos X para la determinación de estructuras superficiales. Este experimento convierte, pues, esta técnica en una alternativa poderosa a la difracción de electrones lentos LEED, de los que se ocuparon el autor y Van Hove en el trabajo antes citado.

Cuando intentamos obtener información sobre la estructura superficial no podemos medir el coeficiente de absorción de los rayos X, pues éstos penetran profundamente en el interior del cristal





Representación esquemática de parte del espectro de absorción del hierro (**arriba**) *y aumento de una región cerca del borde de absorción* (**abajo**)

y, por tanto, la información sobre la superficie que contienen al emerger aparece totalmente diluida con respecto a la del interior del cristal. Lo que uno recoge y analiza son los electrones que se emiten como consecuencia de la absorción del fotón; generalmente, uno mide electrones secundarios que se producen en procesos Auger, es decir, electrones que pueden escapar del átomo al adquirir parte de la energía que se libera cuando el estado vacío, hueco, que había dejado el electrón que absorbe el rayo X es rellenado por otro electrón que se encuentra en un nivel de más energía. Dichos electrones sufren también los procesos de interferencia citados antes en el texto, y como consecuencia, llevan con ellos la información estructural que nos interesa. Además, y éste es el punto clave, la distancia media recorrida por un electrón antes de perder energía, es decir, el recorrido libre medio de dichos electrones, es pequeña: unos 10 angstrom. Los electrones emergentes sólo llevan información sobre la parte del sólido que nos importa, el equivalente de unas cuatro capas atómicas: la superficie.

Usando esta técnica, Bianconi y Bachrach han conseguido determinar la estructura atómica de las superficies 100 y 111 del aluminio. Estas superficies habían sido ya estudiadas antes mediante LEED. El análisis LEED indica que la superficie 100 (los enteros 100 describen la orientación de un plano en el cristal, la superficie es paralela a dicho plano) tiene la misma estructura que la que tendría en el interior del material: no ha habido una reconstrucción. La superficie presenta una contracción del 5-10 por ciento de la distancia interatómica. Es decir, la distancia entre el plano de átomos de la superficie y siguiente es menor que la de dos planos equivalentes en el interior del sólido. El poder resolutivo de LEED es aproximadamente 0,1 angstrom, que es equivalente a un cambio de un 5 por ciento en la longitud de enlace, por lo que existe claramente una ambigüedad en la interpretación de pequeños cambios en el espaciamiento interatómico. Por el contrario, el experimento de A. Bianconi y R. Z. Bachrach prueba definitivamente la no reconstrucción de la cara 100, así como la existencia de una contracción para la cara de valor 0,15 angstrom con un error no superior a 0,05 angstrom.

Estos experimentos confirman la gran potencialidad de SEXAFS como prueba de la estructura superficial. Sería muy interesante un estudio sistemático mediante SEXAFS de las superficies que ya han sido estudiadas por LEED.

Hay que señalar que, al menos en principio, SEXAFS puede ser aplicado al estudio de superficies amorfas y otros casos de interés en los que LEED no puede proporcionar información precisa. Es de esperar que en los próximos años esta técnica SEXAFS se aplique extensivamente al estudio de estructuras superficiales con lo que constituirá un poderoso complemento a la difracción de electrones lentos (LEED) en la determinación de estructuras superficiales. (P. M. E.)

E pluribus unum

Una esperanza (y un prejuicio) permanente de los físicos es que el mundo es tan simple que todas las partículas elementales constituyentes de la materia y cada una de las fuerzas que actúan entre ellas pueden describirse mediante una teoría global única. Durante muchos años una "gran unifica-

ción" ha parecido una posibilidad remota, pero últimamente se han puesto a nuestra disposición nuevos medios para la construcción de una teoría unificada, y ya se han realizado algunos intentos al respecto. A pesar de que las teorías propuestas todavía son provisionales se han recibido con la suficiente seriedad como para planear pruebas experimentales.

Ya hay un cierto grado de simplicidad en la descripción de la naturaleza dada por los físicos. Se considera que la materia está formada por las partículas llamadas fermiones, las cuales interaccionan entre sí mediante el intercambio de partículas de otro tipo, llamadas bosones. Los fermiones son los ladrillos con los que está construido el mundo y los bosones son el cemento que los mantiene unidos. Sin embargo, la descripción presenta una complejidad creciente cuando se van añadiendo detalles. Es necesario admitir dos clases de fermiones como elementales, o dicho de otra forma, como carentes de estructura interna: los leptones, entre los que se sitúa el electrón, y los quarks, que son los supuestos constituyentes del protón, del neutrón y de muchas otras partículas ligadas a ellos. Los bosones son todavía más variados; están asociados a las fuerzas fundamentales de la naturaleza, que son de cuatro clases: fuerte, débil, electromagnética y gravitacional.

Podría concebirse un universo más simple. No existe ninguna necesidad aparente para los dos tipos de fermiones elementales y tampoco es obvio por qué deben existir cuatro tipos de fuerzas. Bastaría con un tipo de cada. Las fuerzas son particularmente desconcertantes por su diversidad: cada fuerza tiene su alcance característico propio, y estos alcances difieren en casi 40 órdenes de magnitud.

Para que una teoría describa todas las partículas y fuerzas, es necesario que de alguna forma se vea que todas esas complicaciones son sólo superficiales. Un primer paso hacia esta meta, que ha guiado todos los esfuerzos posteriores, se realizó en el decenio 1960-1969 con la unificación teórica de las interacciones débiles y electromagnéticas. Ambas fuerzas son muy distintas en sus manifestaciones externas. Por ejemplo, el electromagnetismo tiene un alcance infinito, mientras que el alcance de la fuerza débil es, para todos los propósitos prácticos, de sólo 10^{-15} centímetros. ¿Qué significa, pues, decir que estas fuerzas han sido unificadas? Se dice que se han unificado porque todas las propiedades de ambas fuerzas pueden deducirse de una sola teoría fundamental. A modo de analogía digamos que un hombre es distinto de un cocodrilo, pero las dos especies están unificadas en el sentido en que sus respectivos orígenes pueden explicarse mediante una única teoría, la evolución por selección natural.

La unificación de las interacciones débil y electromagnética puede comprenderse imaginando que las leyes de la naturaleza han evolucionado desde un estado original, más simple, hasta el actual, más complicado. En el estado original, la fuerza débil y la electromagnética eran indistinguibles: tenían la misma fuerza y el mismo infinito alcance, y eran transmitidas mediante cuatro bosones, carentes todos ellos de masa. Las fuerzas presentaban simetría, de suerte que podían intercambiarse libremente; tanto si se aplicaba una fuerza como si se aplicaba otra, el efecto era el mismo. Las diferencias observadas en la actualidad se deben al proceso denominado ruptura espontánea de la simetría, que destruyó su equivalencia en intensidad y

alcance. Además, al romperse la simetría, sólo uno de los bosones, el fotón o cuanto de radiación electromagnética, permaneció con masa nula. Los otros tres bosones, que reciben la denominación general de bosones vectoriales intermedios, adquirieron grandes masas. Al expresar estas masas en unidades de energía, los valores se aproximan a los 100.000 millones de electronvolt, o 100 GeV. (Como comparación indiquemos que la masa de un protón es algo menor que un GeV.)

También se ha desarrollado una teoría para las fuerzas fuertes, las cuales actúan sólo entre quarks (y de un modo distinto entre las partículas formadas por quarks). Actualmente parece que por lo menos hay seis "sabores" de quarks, que han sido denominados arriba (*up*), abajo (*down*), extrañeza (*strangeness*), encanto (*charm*), fondo (*bottom*) y cima (*top*). Cada uno de estos seis tipos de quarks se presenta en tres variedades, a las que se ha dado el caprichoso nombre de "colores". Los colores gobiernan las interacciones o fuerzas fuertes. Los bosones sin masa que transmiten la fuerza fuerte entre los quarks coloreados se denominan gluones, en reconocimiento directo de su papel como material de unión. Se necesitan ocho gluones, siendo cada uno de ellos una combinación diferente de los tres colores primarios de quark. La teoría que describe las interacciones entre quarks coloreados y gluones recibe el nombre de cromodinámica cuántica.

Las grandes unificaciones que están ahora en estudio combinan la teoría unificada de las interacciones débiles y electromagnéticas con la cromodinámica cuántica, y las agrupan en una única estructura mayor. En estas grandes teorías unificadas, tres de las fuerzas básicas de la naturaleza (la fuerte, la débil y la electromagnética) se tratan como manifestaciones distintas de un único fenómeno subyacente. Además, los quarks y los leptones están incorporados a las mismas familias.

En realidad se necesitan tres "generaciones" distintas de fermiones, cada una de ellas formada por dos sabores de quark y dos leptones. La primera generación consiste en los quarks arriba y abajo (cada sabor en los tres colores) y en el electrón y el neutrino electrónico. Estas cuatro partículas (u ocho, si se cuenta cada color por separado) son todo lo que se necesita para explicar la composición de la materia ordinaria. Las restantes partículas aparecen sólo en experimentos de alta energía. En la segunda generación se combinan los quarks extrañeza y encanto con dos leptones más, el muon y el neutrino muónico. La tercera generación incluye los quarks fondo y cima, el leptón conocido por el nombre de tau y su neutrino asociado.

La estructura de una gran teoría unificada sería muy similar a la de la unificación de las fuerzas débiles y electromagnéticas, pero tendría un campo de aplicación mayor. Puede pensarse que en algún estado primitivo las tres fuerzas poseían intensidades idénticas, y que todos los leptones y quarks de cada generación carecían de masa. Las propiedades distintivas de las tres fuerzas se introdujeron debido a la ruptura espontánea de una simetría, lo que también dio masas a la mayoría de fermiones al tiempo que dejaba un fermión en cada generación (el neutrino) con masa nula. En este esquema hay una jerarquía de simetrías rotas. La primera ruptura de simetría fue responsable de la separación entre la fuerza fuerte, por un lado, y la débil y la electromagnética, que todavía permanecían unificadas, por otro. La ruptura de la segunda simetría originó la distinción entre interacciones débiles y electromagnéticas. Naturalmente, la secuencia temporal de sucesos imaginada aquí es del todo artificial: ambas simetrías se rompieron desde el principio.

Si los quarks y los leptones están incluidos en las mismas familias, debe haber algún mecanismo capaz de convertir un tipo de partículas en el otro. En ausencia de una gran unificación, la transformación de un quark en un leptón (o viceversa) está absolutamente prohibida. Un gluon puede cambiar el color de un quark, y un bosón vectorial intermedio puede cambiar el sabor de un quark o de un leptón, pero la frontera entre quarks y leptones es inviolable. Cualquier gran teoría unificada introduce una quinta fuerza básica en la naturaleza, llamada hiperdébil, que es la que gobierna la transición quark-leptón. Asociado a esta nueva fuerza se encuentra otro conjunto de bosones que, como resultado de la ruptura de la simetría, adquieren una enorme masa, que quizá llegue a los 10^{15} GeV. Debido a esta masa casi palpable, los bosones de la fuerza hiperdébil han recibido la denominación de "pelotas de béisbol" vectoriales intermedias.

Otro modo de expresar la esencia de la unificación es indicar que la intensidad de cada una de las fuerzas no es constante, sino que depende de la energía con la que interactúan las partículas. Al aumentar la energía de una interacción, la fuerza débil se hace más intensa y lo mismo, pero a un ritmo menor, le sucede a la fuerza electromagnética. Por otra parte, la fuerza fuerte disminuye de intensidad. A energías suficientemente altas las tres fuerzas han de manifestar la misma intensidad.

La energía a la que se espera que converjan las intensidades de las tres fuerzas es de unos 10^{15} GeV, que coincide con las masas de las pelotas de béisbol vectoriales intermedias. Por ahora no hay esperanzas de alcanzar esta energía en el laboratorio (el mayor acelerador aún ha de alcanzar los 10^3 GeV). Pero hubo un tiempo en la historia del universo en que las partículas que tenían esa energía debieron ser comunes. Estamos aludiendo al primer milisegundo después de la "gran explosión". En ese breve intervalo, las simetrías de las fuerzas básicas habrían sido del todo evidentes, y los quarks y los leptones debieron hallarse en equilibrio térmico, de suerte que un tipo de fermión podía convertirse libremente en el otro. Las simetrías se encuentran obscurecidas hoy porque la temperatura del universo ha bajado un abismo; ha sufrido una especie de congelación.

A pesar de que el régimen energético en donde puede observarse directamente esta unificación se halla fuera de nuestras posibilidades, las teorías unificadoras explican cierto número de fenómenos que son observables. Por ejemplo, constituye un enigma que viene intrigando desde hace tiempo por qué el electrón y el protón tienen cargas eléctricas exactamente de la misma magnitud, aunque de signo opuesto. Si no sucediera así, el mundo sería muy distinto, pero esa razón no convence a nadie. Ahora bien, cuando el electrón y los quarks que integran el protón se agrupan en la misma familia, la cuantificación de la carga eléctrica emerge de forma natural de la teoría. Las masas de los quarks, que parecían bastante arbitrarias, pueden calcularse asimismo en el marco de una gran teoría unificadora; para los quarks más pesados, estos cálculos presentan buena concordancia con los valores deducidos experimentalmente, pero con los quarks más ligeros surgen discrepancias que no se han explicado aún. Una gran unificación también ofrece la posibilidad de explicar cómo ha evolucionado el universo con un exceso de la materia sobre la antimateria.

La predicción más extraordinaria de las teorías unificadas es que la materia es inestable. Si los quarks pueden convertirse en leptones, entonces va a ocurrir lo mismo con las partículas formadas por quarks. La más notable de

ellas, el protón, que en teorías menos globales es absolutamente estable, puede desintegrarse a través de reacciones que dan como productos finales sólo fotones, electrones, muones y neutrinos. Esto implica que los átomos también se desintegran, y que el universo en su totalidad es efímero. No obstante, el tiempo necesario para que el mundo se evapore es tranquilizadoramente largo: la vida media del protón se estima en unos 10^{30} años, lo que es más de 10^{28} veces la edad actual del universo.

El detectar la desintegración del protón sería el modo más drástico y concluyente de confirmar una gran teoría unificada. Ya se ha establecido un límite inferior para la vida del protón, y resulta estar cerca de la vida media predicha. En la actualidad se están diseñando experimentos más sensibles por tres grupos de investigadores. Si la vida media es de 10^{30} años, la observación de 10^{30} protones revelaría una desintegración por año. Los experimentos proyectados requieren el control de más de 10.000 toneladas de agua mediante instrumentos capaces de registrar las desintegraciones protónicas y distinguirlas de los numerosos acontecimientos de otro tipo.

Si alguno de los actuales esquemas para la unificación de las fuerzas fuertes, débiles y electromagnéticas resultase ser correcto, se habría obtenido, sin duda, una teoría magnífica, pero aún no sería la teoría definitiva de la naturaleza. Otra fuente de insatisfacción es que se necesitan tres generaciones separadas de fermiones, dos de las cuales parecen ser superfluas. Hay otro problema que puede resultar más incómodo: una de las fuerzas básicas, la gravitación, se ha dejado completamente de lado. La energía característica de las grandes teorías unificadas no es mucho menor que una cantidad conocida por el nombre de masa de Planck, que es la masa para la que las interacciones gravitatorias entre las partículas empiezan a tener una influencia importante en su comportamiento. La masa de Planck tiene un valor, en unidades de energía, de unos 10^{19} GeV, frente a la masa de 10^{15} GeV para la que se unifican las restantes fuerzas. La similitud de estos grandes números puede no ser una coincidencia; de hecho una teoría de las interacciones entre partículas elementales que trate masas tan grandes tendría que incluir los efectos gravitacionales si se buscan resultados precisos. Se han realizado ya ciertos pasos preliminares hacia ese fin. Estos implican los conceptos de supergravedad y supersimetría, donde los bosones y los fermiones se encuadran en las mismas familias.

Teoría de nudos

Un rollo de cordel anudado, con los extremos unidos, proporciona un buen modelo material de nudo matemático abstracto. La teoría de nudos se propone analizar las diferentes formas posibles de trazar en el espacio curvas anudadas

Lee Neuwirth

Los nudos no sólo sirven de ayuda a los marineros y de reto a los boy-scouts; son además tema de investigación matemática seria. En matemáticas, los nudos tienen una definición abstracta, a saber, son curvas unidimensionales situadas en el espacio tridimensional ordinario, que comienzan y terminan en un mismo punto, y que no se cortan a sí mismas. A pesar de ello, la teoría de nudos, que se propone clasificar y analizar las distintas formas de trazar tales curvas no carece totalmente de relación con los cotes, vueltas mordidas, margaritas y demás nudos marineros: un ovillo de cordel cuyos extremos estén empalmados –lo que hará imposible desenredarlo– constituye un excelente ejemplo material del objeto matemático abstracto. Además, algunos de los más profundos problemas de esta rica y variada teoría se presentan justamente al intentar responder las mismas cuestiones que nos plantearíamos para un cordel anudado. ¿Está realmente anudado? Haciendo un segundo nudo, ¿se podrá deshacer el primero? En particular, ¿cuándo son equivalentes dos nudos?

Esta última pregunta suscita el problema fundamental de la teoría de nudos. ¿Cómo podremos distinguir nudos diferentes? Para mejor percibir la esencia del problema, quizá convendrá reflexionar un momento sobre él, no matemática, sino intuitivamente. Los matemáticos son libres de construir teorías a su gusto y medida; pero de ceñirse a la realidad material, la teoría poseería además un evidente valor práctico. Fijémonos en las curvas anudadas que se muestran en la ilustración de la página opuesta. Si se realizasen con cuerda, casi todo el mundo opinaría que las dos figuras superiores están anudadas de diferentes formas. La figura circular está, evidentemente, sin anudar (de hecho, en la jerga de la especialidad, este nudo se llama "nudo trivial"). En cambio, no parece existir ninguna forma de desenredar sin cortarla la figura adyacente al círculo. Por otra parte, aunque las figuras de los restantes pares han sido dibujadas de distinta forma, no parecen existir diferencias esenciales en el procedimiento de anudarlas. Las figuras de la parte inferior, que únicamente se diferencian en la posición del lazo de color, parecen ser equivalentes, lo mismo que la figura de lados rectos y la adyacente a ella, totalmente lisa, y así sucesivamente. ¿De qué manera podrían formalizarse juicios intuitivos como los anteriores? La teoría de nudos busca proporcionar modelos y métodos para demostrar que las dos curvas del par superior son diferentes, despreciando en cambio las "superficiales" diferencias entre las dos curvas de los demás pares.

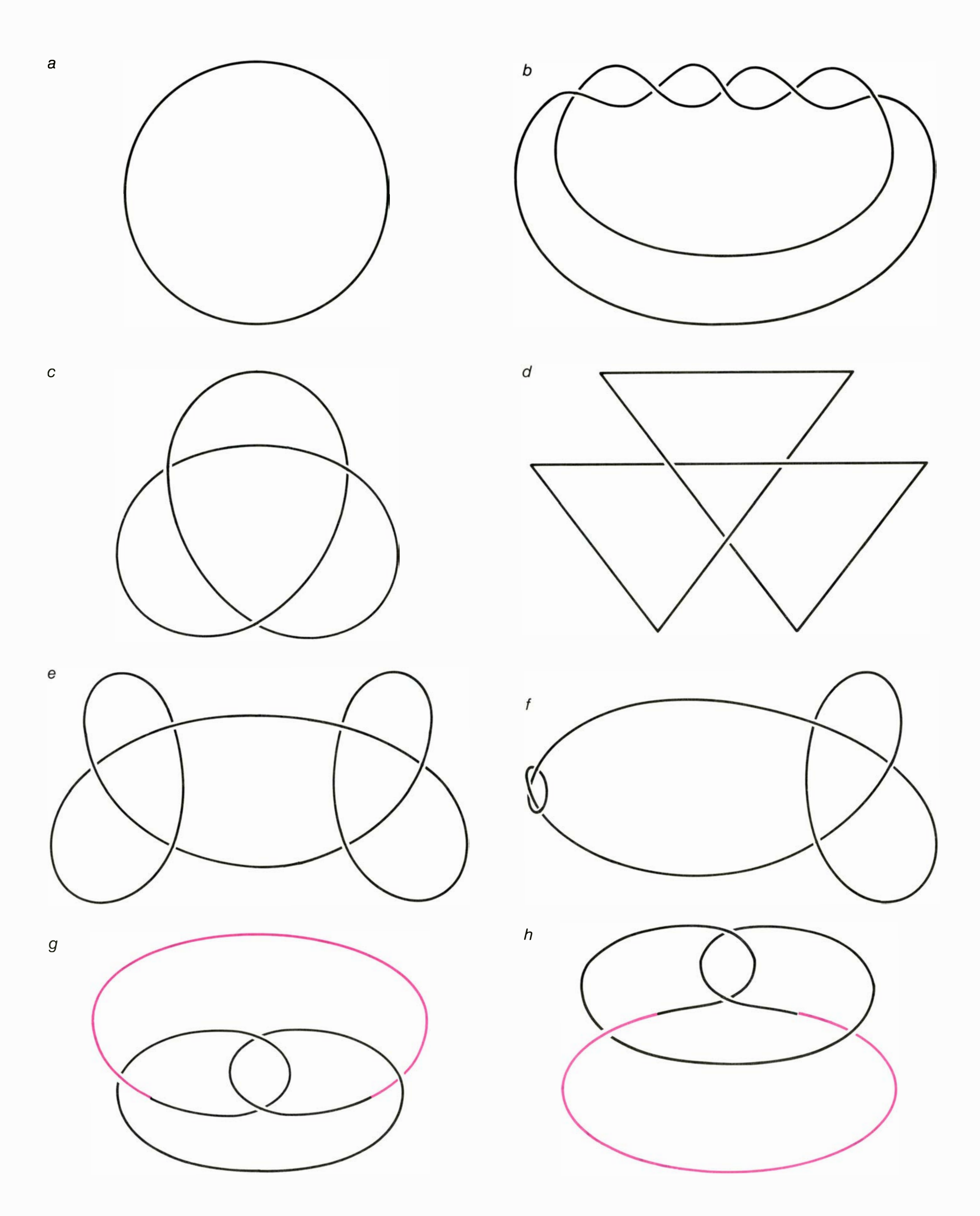
La parcela de la matemática que hace caso omiso de las cualidades de "lisura", tamaño e, incluso, de forma es la rama de la geometría conocida por el nombre de topología. La topología se interesa por aquellas propiedades que permanecen inalterables al deformar arbitrariamente los espacios. Uno de los métodos que aplican los topólogos en la investigación de tales propiedades consiste en sumergir un objeto en un espacio y ver cómo se comporta la parte desocupada del espacio. Para comprender cómo surgen los nudos en investigaciones de este tipo es esencial darse cuenta de que la propiedad de "anudación" no es intrínseca de la curva, sino de la inmersión de la curva, es decir, de la forma en que la curva se encuentra alojada en el espacio tridimensional. Un insecto imaginario que se arrastrase por el espacio unidimensional de una curva anudada no sabría reconocer que la curva estaba anudada; desde el punto de vista del insecto, la curva sería indistinguible de un círculo. (En realidad, una curva unidimensional sólo puede estar anudada en el espacio tridimensional. El espacio de dimensión dos no es lo bastante holgado como para anudar la curva; y el de dimensión cuatro lo es demasiado, es decir, toda curva razonable inmersa en él puede desanudarse.)

Así pues, los topólogos consideran que la teoría de nudos es un modelo bastante sencillo de los problemas llamados "de colocación" o "alojamiento": ¿Cuáles son las distintas formas en que puede sumergirse una circunferencia en el espacio tridimensional? El estudio de los nudos y de los espacios en que están sumergidos es una parcela relativamente reciente de las matemáticas, datando su origen de principios de siglo; no obstante, mucho se ha penetrado en ella. Antes de entrar en la descripción de algunos métodos encaminados a probar que dos nudos son distintos, es, sin embargo, necesario decidir cuándo se los podrá considerar iguales.

Nudos equivalentes

Existen definiciones matemáticamente rigurosas para precisar qué nudos son equivalentes; y aunque podríamos presentarlas aquí, exigirían buena cantidad de conceptos matemáticos previos. Afortunadamente, mediante analogías con el mundo físico podemos conseguir casi los mismos resultados. Para determinar si dos nudos son equivalentes, empezaremos imaginándolos como cuerdas anudadas. Para ser más exactos, tomemos un modelo de cada nudo, que obtendremos engrosando ligeramente el nudo, como si se hallase alojado en un tubo tridimensional sólido y flexible. Se dirá que dos nudos son equivalentes cuando el modelo correspondiente a uno de ellos pueda deformarse –estirándolo, contrayéndolo o retorciéndolo– hasta alcanzar la forma del otro, sin romper el tubo ni hacerlo pasar a través de sí mismo. (Ciertos nudos "extravagantes", como el mostrado en la ilustración inferior de la página 54, no pueden ser alojados en un tubo tridimensional; al no corresponder su estudio a la teoría clásica de nudos, no me referiré a ellos.)

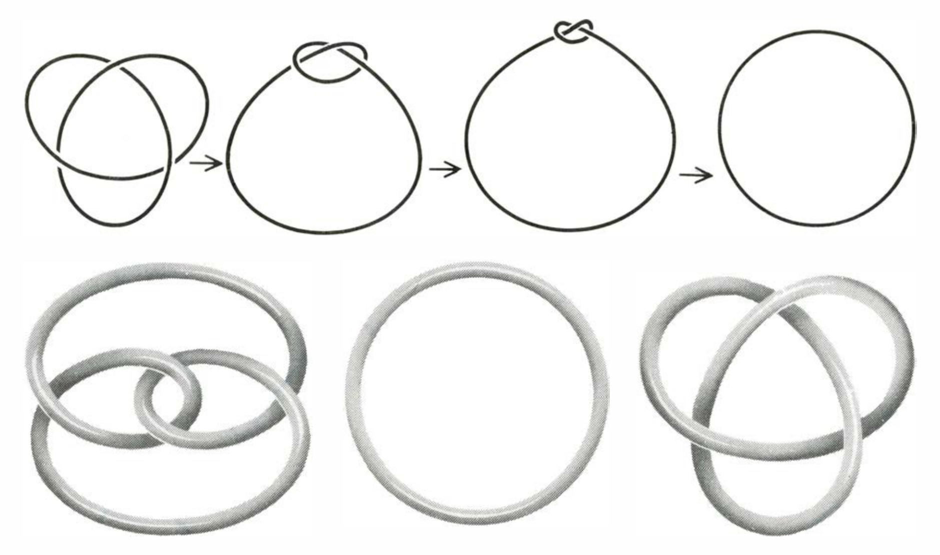
Para demostrar que dos nudos son



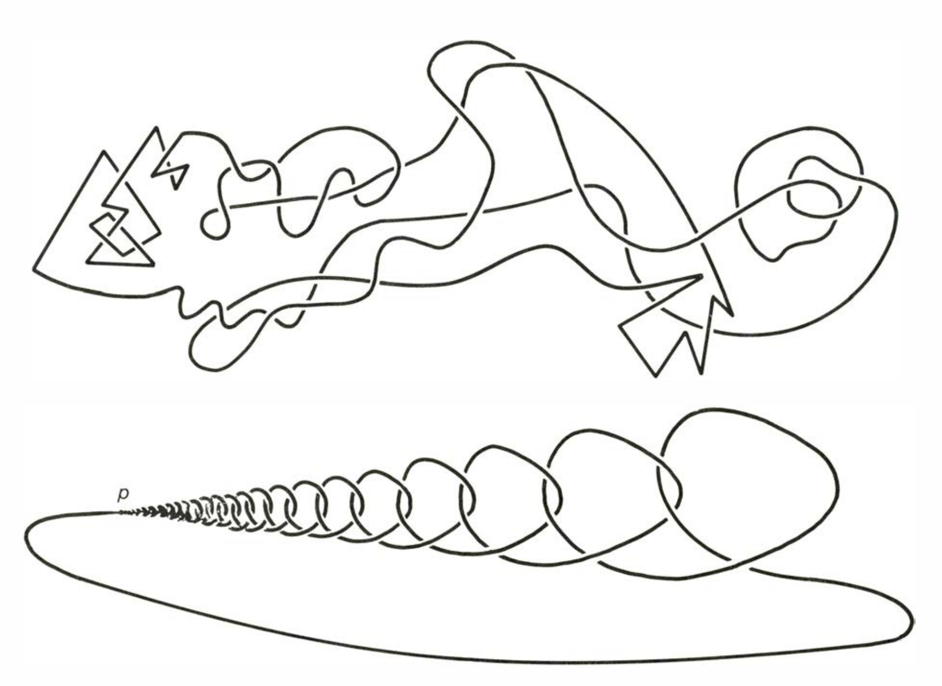
UN NUDO es una curva unidimensional trazada en el espacio de dimensión tres, de forma que se inicie y termine en un mismo punto y no se corte a sí misma. Todas las figuras aquí mostradas representan nudos, incluso la circular, *a*, llamada nudo trivial o nudo desanudado, así como la figura *d*, que tiene los lados rectos. La teoría de nudos es parte de la topología, rama de la geometría que estudia las propiedades de los espacios que permanecen invariables al deformarse éstos. Tanto los nudos *c* y *d*, que únicamente se distinguen por la forma de sus lazadas, como los nudos *e* y *f*, que difieren sólo en el tamaño de sus lazos, son considerados equivalentes, y lo mismo para los nudos *g* y *h*, que se distinguen exclusivamente por la posición de un lazo (*en color*). Por otra parte, el círculo sin anudar parece ser distinto de la curva anudada *b* que se muestra a su lado. El problema de probar que tales nudos, o curvas anudadas, son verdaderamente distintas entre sí constituye una de las principales tareas a dilucidar por la moderna teoría de nudos.

equivalentes bastará deformar el modelo tubular de uno de ellos hasta que llegue a coincidir con el otro modelo. (Las curvas anudadas unidimensionales no pueden tratarse de esta forma, porque los nudos no triviales pueden siempre escurrirse de ellas, como indica la ilustración superior de esta misma página.) Refiriéndonos otra vez a los pares de curvas anudadas intuitivamente equivalentes que se ven en la ilustración de la página anterior, es obvio que sus modelos tubulares pueden efectivamente hacerse coincidir. Por otra parte, pronto resulta evidente que no hay forma ninguna de retorcer el modelo circular hasta darle la forma del nudo que vemos a su lado. El hecho de que fracasen todos los intentos no basta, sin embargo, para demostrar que estos dos nudos (u otros dos cualesquiera) son realmente distintos, porque siempre queda la duda de que mediante algún hábil truco se consiga deformar uno de los modelos hasta convertirlo en el otro. Por tanto, para demostrar que dos nudos son diferentes hay que descubrir alguna propiedad que los distinga. Las propiedades de este tipo se llaman "invariantes".

Los invariantes son propiedades características de los nudos que permanecen inalterables al deformar el modelo tubular correspondiente. Por ejemplo, el número mínimo de puntos de cruce necesarios para dibujar el nudo es un invariante, que suele utilizarse como criterio de organización de tablas de nudos. No obstante, el número de cruces es, en ciertos aspectos, un invariante poco satisfactorio. Para empezar, no siempre es evidente que el dibujo del nudo haya sido trazado con el número mínimo posible de cruces. No es, pues, fácil calcular tal número. Además, muchos nudos diferentes tienen el mismo número de cruces; este número no es, por tanto, un invariante muy potente. Dicho de otra forma, el número de cruces no basta para diferenciar la mayoría de los nudos. Finalmente, dado que un único número no podrá contener mucha información sobre la estructura del nudo, parece que este invariante no facilitará gran cosa la percepción de otras propiedades de los nudos.



UN NUDO NO TRIVIAL de una curva unidimensional puede siempre "escurrirse" deformándolo hasta un punto, como se muestra en la parte superior. No siendo esto cierto para cuerdas anudadas, es conveniente imaginar modelos de curvas anudadas como los mostrados abajo. Para construirlos se ha engrosado ligeramente cada curva, con lo que éstas parecen hallarse alojadas dentro de un tubo tridimensional sólido y flexible. Dos nudos podrán considerarse equivalentes cuando el modelo tubular de cada uno pueda deformarse hasta conseguir el otro sin romper el tubo ni hacer que se atraviese.



UNA CURVA IRREGULAR ANUDADA, como la que se ve arriba, es considerada nudo aceptable, porque puede alojarse en un tubo sólido flexible y tridimensional. En cambio, la curva mostrada abajo no puede alojarse en tubos así, porque al entrecruzarse sus lazos convergen hacia un único punto, *p*. Las curvas de este tipo se llaman "extravagantes" y su estudio no corresponde a la teoría de nudos.

En la actualidad, pese a existir varias refinadas técnicas que podrían servir en principio para determinar si dos nudos cualesquiera son o no equivalentes, no se dispone de procedimientos prácticos para llevar a cabo tal comprobación. Dicho de otra forma, no existen invariantes de deducción sencilla que caractericen totalmente cada nudo. Sí existen, sin embargo, varios invariantes que casi permiten obtener una clasificación completa. Uno de ellos, que describiré con cierto detalle, se conoce por el nombre de grupo nodal. Hablando en términos generales, el grupo nodal describe las distintas formas en que es posible cruzar el espacio tridimensional sin tropezarse con un nudo inmerso en él.

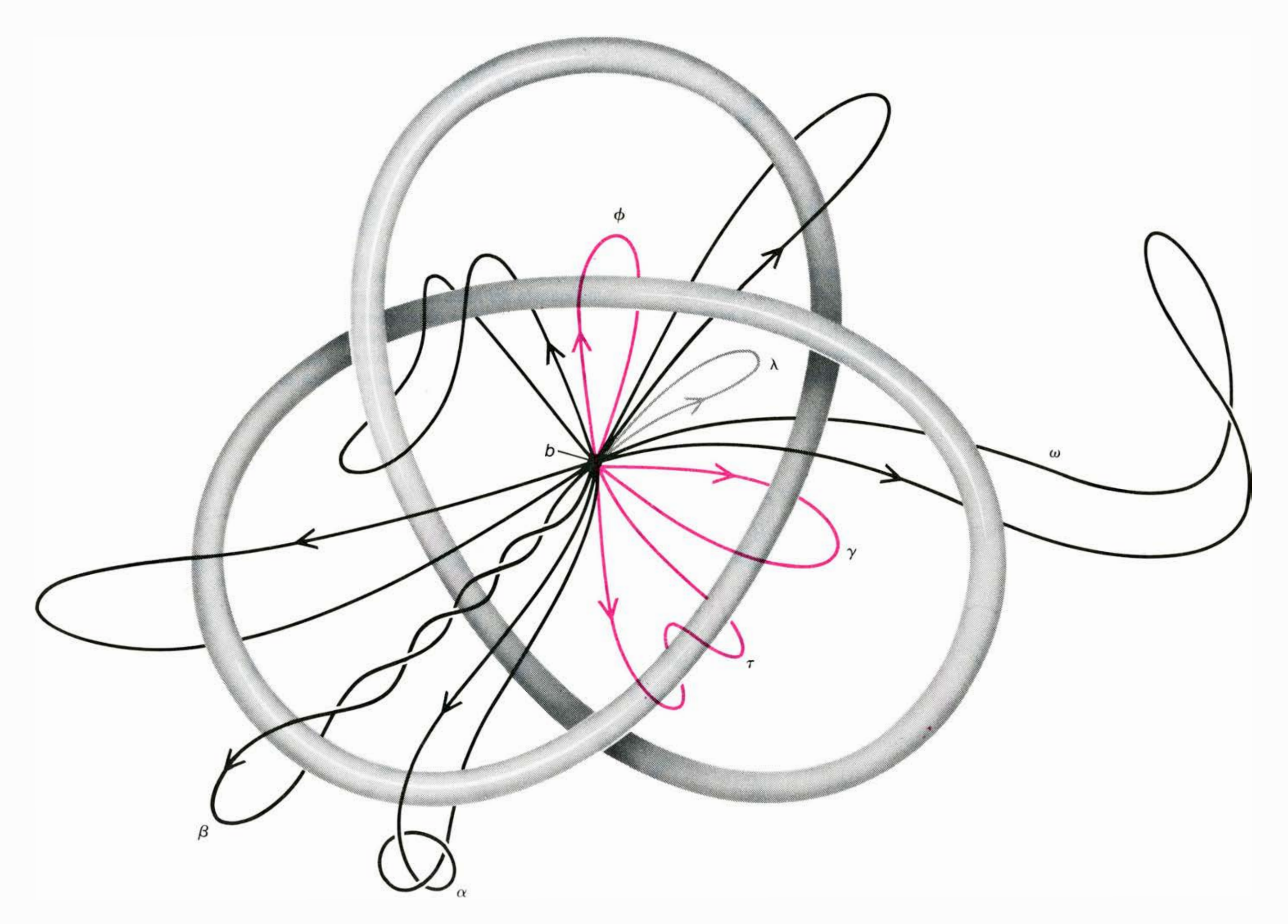
Construcción del grupo nodal

El espacio que rodea un nudo inmerso, y las trayectorias o caminos que lo surcan, pueden ser descritos muy adecuadamente por medios geométricos, pues la geometría estudia problemas que admitan representación espacial. Hace unos 80 años, Henri Poincaré descubrió que propiedades importantes de esta configuración geométrica podrían describirse también por medios algebraicos. El álgebra investiga problemas con representación simbólica. El grupo nodal es solamente uno de los invariantes algebraicos de los nudos que se han ideado desde entonces. La imposición de estructura algebraica sobre un espacio intrínsecamente geométrico suele presentar dificultades formidables que no siempre vale la pena afrontar. No obstante, en el caso de los nudos, ha ocurrido que la mayoría de los problemas resultan más fáciles de resolver por

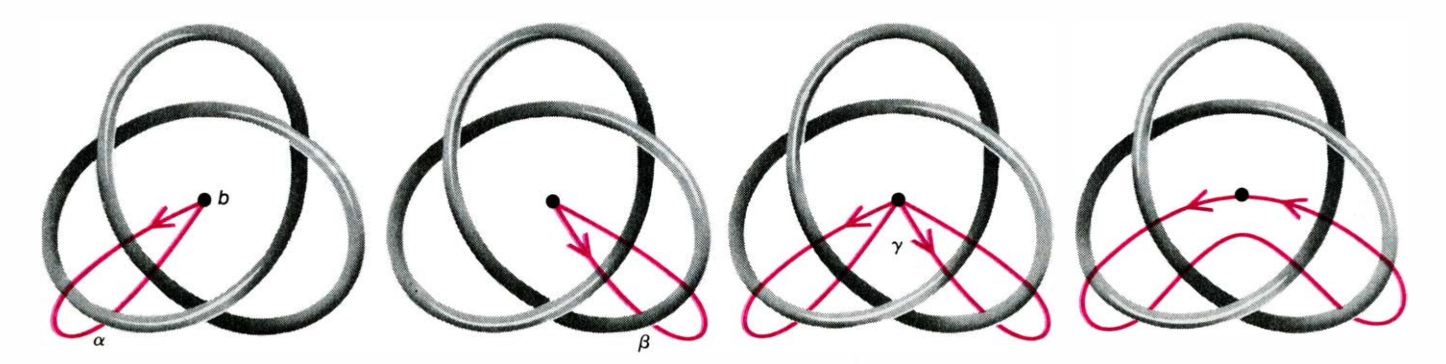
manipulación de símbolos. En realidad, gran parte de la belleza y profundidad de la teoría de nudos se debe al juego contrapuntístico entre geometría y álgebra, dos de las más antiguas ramas de la matemática, que se presentan aquí bajo nuevas formas, revitalizadas con ideas nuevas.

Para comprender cómo a un nudo inmerso en un espacio tridimensional se le puede asociar una estructura algebraica –la de grupo– fijémonos en el modelo de nudo en trébol de la ilustración inferior. La región de espacio tridimensional no perteneciente al nudo inmerso se llama complemento del nudo, y, por comodidad, llamaré *S* al complemento de este tubo anudado y tomaré en *S* un punto fijo *b*. (La estructura del grupo nodal resultará ser independiente de la elección de *b*.) La ilustración muestra varios elementos de Ω, es decir, del conjunto de todos los caminos unidimensionales que parten de *b*, recorren *S* (evitando por consiguiente tropezar con el tubo anudado) y finalmente retornan a *b*. Obsérvese que el conjunto Ω está formado por lazos (ciclos) cerrados y con un sentido de recorrido, contenidos en *S*, y que pueden o no interceptarse a sí mismos. Entre ellos se cuenta el lazo constante o trivial, *e*, que es una trayectoria que no abandona el punto base *b*. Aunque, como es evidente, el conjunto Ω es muy grande, existe un procedimiento de clasificación de sus elementos que lo hace mucho más manejable.

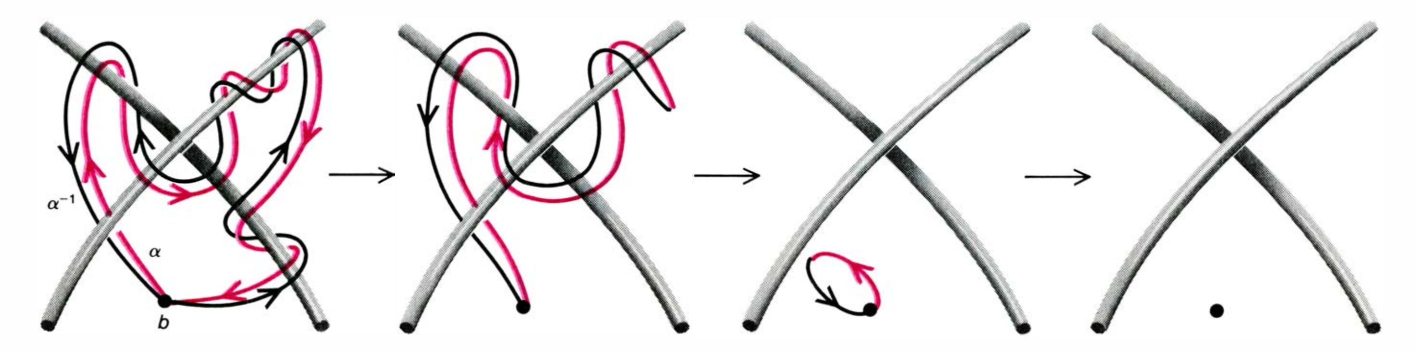
El conjunto Ω puede descomponerse en clases de caminos equivalentes, declarando ser dos caminos α y β equivalentes cuando cada uno de ellos puede deformarse hasta coincidir con el otro. Durante el proceso de deformación, α podrá ser estirado y encogido, e incluso podrá hacerse que se cruce consigo mismo, pero ni su punto inicial ni su extremo pueden sufrir alteración; tampoco es lícito romperlo ni hacerlo salir de la región *S*. (Dicho de otra forma, α no puede entrar en contacto con ningún segmento del nudo.) Por ejemplo, entre los elementos de Ω representados en la ilustración, el camino λ puede contraerse hasta el punto base *b* (y es equivalente, por tanto, al camino constante *e*); en cambio, el camino β que forma espira alrededor de un segmento del nudo no puede encogerse hasta tan atrás. No obstante, el camino β puede destorcerse y el "nudo" del camino α puede ser eliminado haciendo que el camino se atraviese a sí mismo; resulta así que



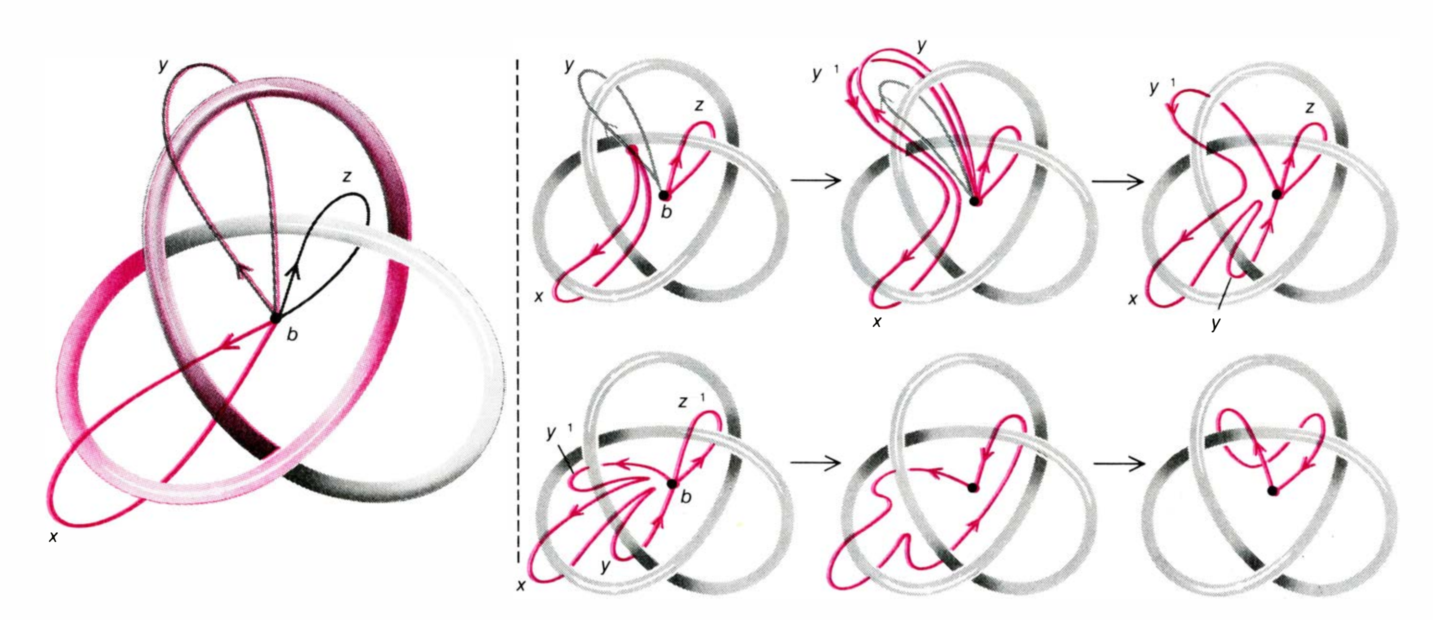
LA ANUDACION no es propiedad de la curva, sino de la forma en que ésta se encuentra inmersa en el espacio tridimensional. Así pues, un método para distinguir un tipo concreto de nudo consiste en caracterizar de qué formas puede cruzarse el espacio sin tropezar con el nudo inmerso. Se muestran aquí algunos de los caminos cerrados que cruzan el espacio que rodea al nudo en hoja de trébol, partiendo de un punto base *b*, fijo y arbitrario. (En teoría de nudos, la región de espacio tridimensional exterior a un nudo inmerso, llamada complemento del nudo, se supone siempre no ocupada por otros objetos.) No todos los caminos que parten de *b* representan rutas topológicamente diferentes a través del espacio complementario del nudo, y por tanto, si existe algún procedimiento para deformar un camino en otro, entonces ambos se consideran equivalentes. En tales deformaciones, llamadas homotopías, es lícito estirar o encoger arbitrariamente los caminos, e incluso cruzarlos a través de sí mismos, pero sus puntos extremos han de permanecer quietos; tampoco es lícito hacer que atraviesen un segmento del nudo. Todos los caminos trazados en negro son homotópicos; por ejemplo, el "nudo" del camino α puede eliminarse haciendo que el camino se atraviese a sí mismo; el camino β puede destrenzarse. Los caminos de color no son homotópicos entre sí (ni tampoco a los negros); por ejemplo, los caminos γ y ω rodean el mismo segmento del nudo, pero en sentidos diferentes, el camino τ rodea al mismo segmento de nudo que el ω, y en el mismo sentido, pero da una vuelta de más, y el camino Φ pasa alrededor de un segmento distinto. El camino λ puede retraerse hasta el punto base *b*, por lo que es homotópico al camino constante *e*, que nunca abandona *b*. La introducción de la relación de equivalencia definida por la homotopía transforma el conjunto de caminos que surcan el complemento de un nudo en una colección de clases de caminos homotópicos. La clase de caminos homotópicos a un cierto camino determinado α se denota [α].



COLECCION DE CLASES DE HOMOTOPIA de caminos trazados en el complemento del nudo en hoja de trébol, que puede manejarse algebraicamente, es decir, simbólicamente, imponiendo sobre el conjunto de clases la forma de multiplicación mostrada aquí. El producto $[\alpha]\,[\beta]$ de las clases representadas por α y β es, por definición, la clase $[\gamma]$, siendo γ el camino $\alpha\beta$ que comienza en *b*, recorre primero α y seguidamente β, terminando en *b*. Otro de los caminos perteneciente a $[\gamma]$ se muestra en la figura de más a la derecha. No es difícil comprender que este producto es asociativo, por lo que $([\alpha]\,[\beta])\,[\tau]$ es igual a $[\alpha]\,([\beta]\,[\tau])$ para cualesquiera clases $[\alpha]$, $[\beta]$ y $[\tau]$. El producto no es conmutativo, o sea, $[\alpha]\,[\beta]$ no es forzosamente igual que $[\beta]\,[\alpha]$. La clase del camino constante $[e]$ actúa como identidad: para toda clase $[\alpha]$, se verifica que $[e]\,[\alpha]$ e igual a $[\alpha]\,[e]$, y este producto, igual $[\alpha]$.



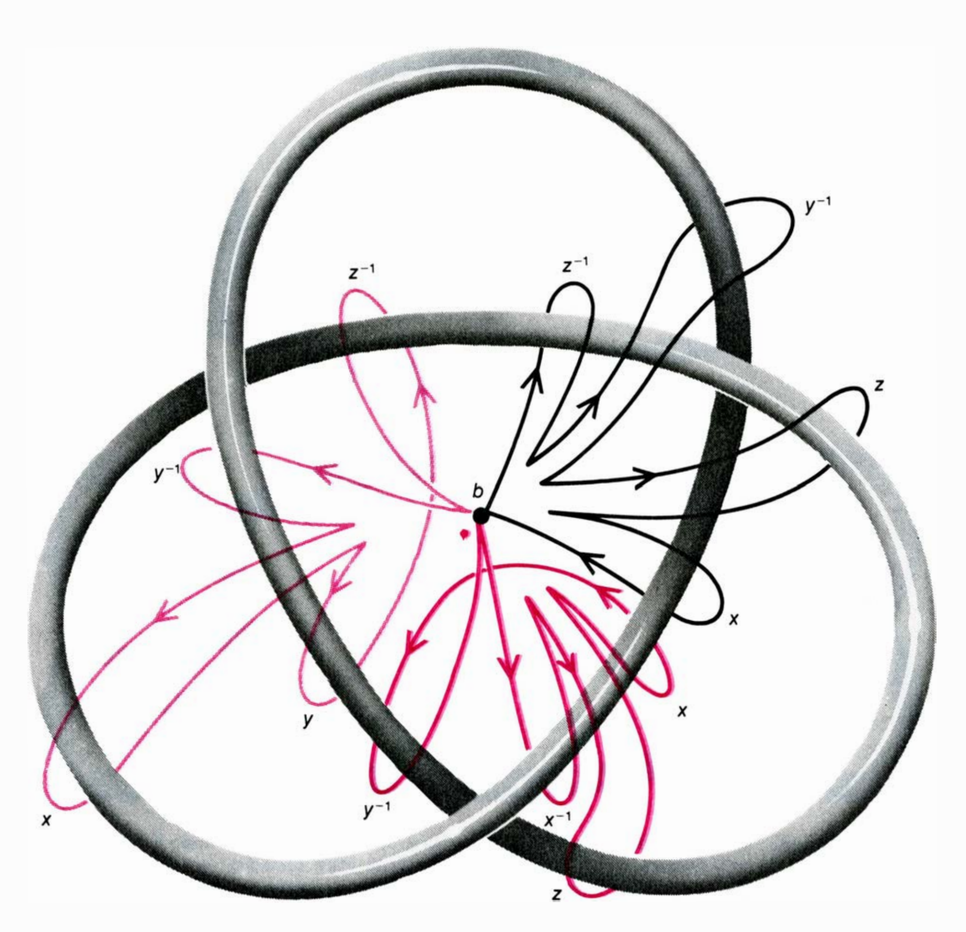
INVERSA DE UNA CLASE DE HOMOTOPIA, que es por definición la clase $[\alpha]^{-1}$ formada por todos los caminos homótopos a α^{-1} (*en negro*), es decir, α recorrido en sentido contrario. Como aquí se ve, $\alpha\alpha^{-1}$ es homotópico a *e*, y por tanto, $[\alpha]\,[\alpha]^{-1}$ que es igual a $[\alpha]^{-1}\,[\alpha]$ es igual a $[e]$. En álgebra, todo conjunto dotado de un producto asociativo con elemento identidad, y donde todo elemento posea inverso, se denomina grupo. El grupo cuyos elementos son las clases de homotopía de caminos que surcan el complemento del nudo se llama grupo nodal. No es difícil ver que al deformar el nudo su grupo permanece invariable. Dicho de otra forma, el grupo es un invariante característico del nudo. Además, ocurre que muchos nudos (aunque no todos) son distinguibles por su grupo; de ello se deduce que el grupo nodal es un poderoso instrumento para diferenciar unos nudos de otros.



EL GRUPO NODAL ES CALCULABLE, lo que significa que siempre es posible construir una presentación finita, explícita, del grupo. Al construir la presentación del grupo nodal correspondiente a la hoja de trébol mostrada a la izquierda, se observaría, por ejemplo, que los puntos de cruce del nudo lo dividen en tres segmentos, que van desde cada paso inferior hasta el siguiente. Eligiendo entonces un punto arbitrario *b* en el complemento del nudo, se pueden trazar tres caminos *x, y, z,* cada uno de los cuales rodea un segmento distinto del nudo. Las clases de homotopía $[x]$, $[y]$, $[z]$ se llaman generadores del grupo nodal, porque cualquier elemento del grupo es expresable como producto de estas tres clases y sus inversos. Además, en los puntos de cruce del nudo, los tres segmentos enlazan de distintas formas, definiendo así relaciones entre los generadores. Al expresar la relación correspondiente al cruce superior izquierdo del nudo, pongamos por caso, se debe observar que llevando los extremos del camino *x* a lo largo del camino *y* se origina un camino homotópico a *z*. No obstante, como se muestra en la sucesión superior, a la derecha, tal camino es también homotópico a $y^{-1}\,x\,y$. Por consiguiente, en este cruce $[y]^{-1}\,[x]\,[y]$ es igual a $[z]$, o como se expresa abajo, a la derecha, $[y]^{-1}\,[x]\,[y]\,[z]^{-1}$ es igual a $[e]$. Esta igualdad se llama relación del grupo nodal. Las relaciones entre los generadores correspondientes a los otros dos puntos de cruce se deducen igual.

ambos caminos α y β son equivalentes. Los procesos de deformación de este tipo se llaman homotopías; caminos como α y β, que únicamente difieren en una homotopía, se denominan homotópicos (u homótopos). En la ilustración, todos los caminos trazados en negro son homotópicos entre sí, mientras que ninguno de los dibujados en color es homotópico a ninguno de los demás, sea negro o de color.

La clase de todos los caminos homotópicos a un determinado camino α se denota $[\alpha]$. No obstante, debería ser evidente que la composición de la clase $[\alpha]$ no depende de la elección de α, en el sentido de que si α es homótopo a β, las clases $[\alpha]$ y $[\beta]$ serán idénticas. De esta forma, el conjunto Ω de todos los caminos ha sido descompuesto en una colección de clases de homotopía claramente definidas. Además, es posible definir de forma muy natural una multiplicación de clases: el producto $[\alpha][\beta]$ de dos clases de caminos de Ω es igual a la clase $[\gamma]$, siendo γ el camino $\alpha\beta$ resultante de componer α y β, es decir, un camino que partiendo de b recorre α hasta regresar a b y después prosigue a lo largo de β hasta regresar por segunda vez a b. No es difícil ver que la clase $[\gamma]$, que también podemos llamar $[\alpha\beta]$, está bien definida, en el sentido de que los caminos que la forman no cambian si en lugar de los caminos α y β se toman otros homótopos a ellos.



RELACIONES DEL GRUPO NODAL causadas por los tres puntos de cruce del nudo de hoja de trébol, a saber: $[y]^{-1}[x][y][z]^{-1} = [e]$ (*color claro*), $[z]^{-1}[y]^{-1}[z][x] = [e]$ (*en negro*) y $[x]^{-1}[z][x][y]^{-1} = [e]$ (*color obscuro*). De estas tres relaciones puede deducirse toda identidad que ligue los elementos del grupo nodal. Así pues, como todo elemento del grupo puede expresarse mediante los generadores y sus inversos, la lista de generadores y relaciones del nudo de hoja de trébol describe totalmente el correspondiente grupo. Existen otros muchos algoritmos para deducir representaciones, o sea, listas de generadores y relaciones entre ellos, que describen explícitamente el grupo de un nudo concreto.

El grupo nodal

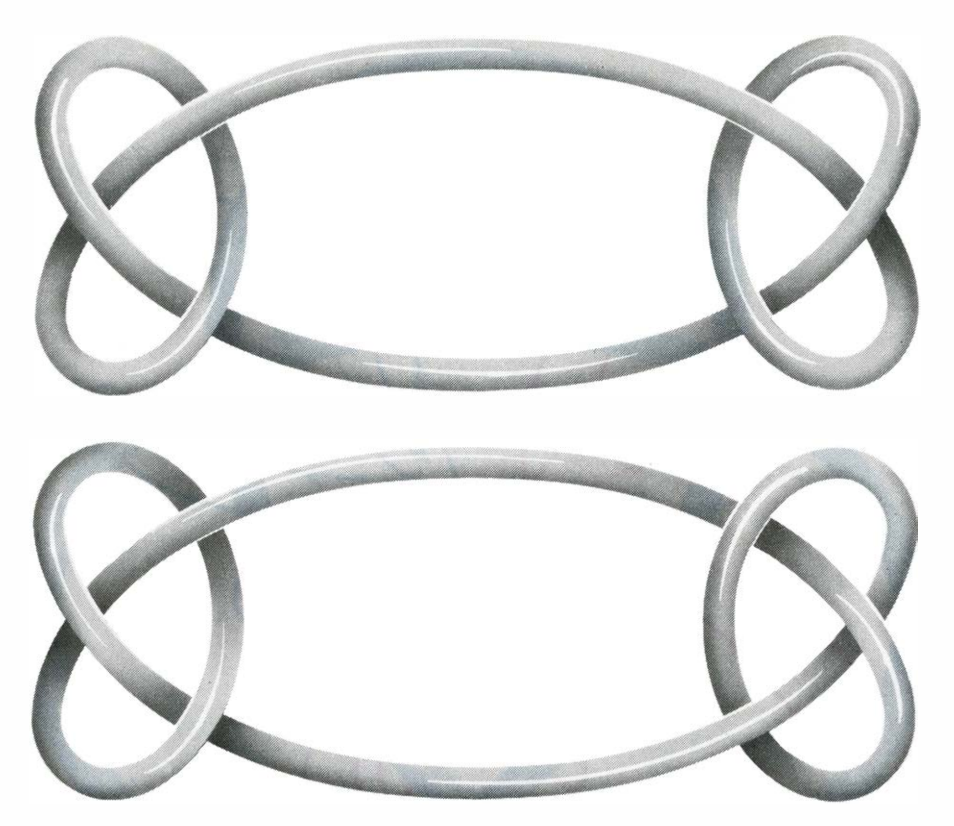
Hasta ahora la descripción del conjunto de caminos que surcan el complemento de un nudo inmerso ha sido de naturaleza geométrica. Al imponer sobre las clases de homotopía de tales caminos una regla de multiplicación comienza a emerger una estructura algebraica, un procedimiento para manejar caminos por métodos simbólicos. Ciertas observaciones concernientes a esta estructura nos ayudarán a identificarla. Ante todo, observemos que la multiplicación de clases es una operación asociativa: cualesquiera que sean los caminos α, β y γ de Ω, $([\alpha][\beta])[\gamma]$ es igual a $[\alpha]([\beta][\gamma])$.

Por otra parte, experimentando un poco se ve pronto que $\alpha\beta$ no suele ser homotópico a $\beta\alpha$. Por consiguiente, la multiplicación de clases de homotopía no es conmutativa, es decir, $[\alpha]\cdot[\beta]$ no suele coincidir con $[\beta][\alpha]$, a menos que $[\alpha]$ o $[\beta]$ sean iguales a la clase del camino trivial $[e]$. La clase $[e]$ actúa como elemento identidad de la multiplicación de clases de caminos de Ω, (lo mismo que actúa el número 1 respecto de la multiplicación de números enteros), es decir, que para toda clase $[\alpha]$, $[\alpha][e]$ es igual a $[e][\alpha]$ y ésta, igual a $[\alpha]$. Por fin, es evidente que para todo elemento $[\alpha]$ existe un elemento inverso $[\alpha]^{-1}$, tal que $[\alpha][\alpha]^{-1}$ es igual a $[\alpha]^{-1}[\alpha]$, que es igual a $[e]$. (Como se muestra en la ilustración central de la página anterior, $[\alpha]^{-1}$ es la clase de caminos homotópicos a α^{-1}, es decir, del camino de Ω formado por α, recorrido en sentido contrario.)

Un conjunto cualquiera, dotado de una operación asociativa definida para todo par de sus elementos, de forma que exista elemento identidad (o neutro) para la operación, y que cada elemento del conjunto posea inverso respecto de la operación, es un objeto algebraico llamado grupo. No es difícil ver que el grupo nodal –a saber, su colección de clases de homotopía de caminos, dotada de la multiplicación arriba impuesta– es un verdadero invariante del nudo. Pensemos en lo que ocurre con el grupo de un nudo al deformar éste –o un tubo anudado– en otro nudo equivalente. En el curso de la deformación, todos los caminos contenidos en el complemento del nudo inicial se convertirán en caminos contenidos en el complemento del segundo nudo. En particular, todo par de caminos homotópicos contenidos en el complemento del primer nudo se transforma en un par de caminos homotópicos en el complemento del segundo. Puesto que lo mismo es cierto para la deformación del segundo nudo en el primero, las clases de homotopía y, por tanto, los grupos correspondientes a ambos nudos serán equivalentes. Con un poco de esfuerzo, estas observaciones pueden transformarse en demostración rigurosa de que el grupo de un nudo es auténticamente un invariante del nudo.

Presentaciones del grupo nodal

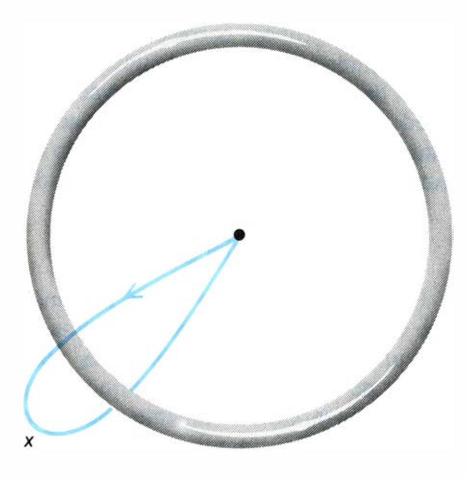
El grupo nodal es un objeto matemático de difícil manejo, pues está formado por un número infinito de clases abstractas de equivalencia de caminos. Así que este invariante solamente podrá ser utilizado para distinguir unos nudos de otros cuando se disponga de un procedimiento para describirlo explícitamente. De hecho, siempre es posible calcular el grupo nodal, es decir, siempre es posible construir una lista de objetos concretos que presenten o describan completamente al grupo. Tal lista estará formada



NO TODOS LOS NUDOS quedan caracterizados por su grupo. Por ejemplo, el nudo llano que se ve arriba y el nudo de rizo mostrado abajo tienen grupos idénticos. Para demostrar que estos nudos son en realidad distintos hay que aplicar técnicas más refinadas. Siempre puede calcularse el grupo nodal.

por cierto número de elementos del grupo, llamados generadores, y por cierto número de ecuaciones, llamadas relaciones.

Para darnos cuenta de cómo deducir una lista de este tipo fijémonos en el nudo que se muestra a la izquierda de la ilustración inferior de la página 60. Los tres puntos de cruce de nudo lo dividen en tres segmentos, tomados desde cada paso inferior hasta el siguiente. (En la ilustración, estos segmentos se han trazado con líneas de distinto tono.) Se ha



NUDO TRIVIAL, o desanudado. Es el que tiene el grupo nodal más simple: el grupo presentado por un único generador [*x*], sin relaciones. No hay otros nudos que sean descritos por este grupo; por tanto, para determinar si cierto nudo está realmente anudado sería suficiente (aunque no siempre es ello fácil) demostrar que su grupo está formado solamente por potencias de [α] y de $[\alpha]^{-1}$.

elegido un punto base fijo en el complemento del nudo, y desde este punto fijo *b* se han trazado caminos, uno para cada segmento del nudo, que rodean el segmento y retornan a *b*. Los tres caminos *x, y, z* obtenidos de esta forma originan las tres clases de homotopía [*x*], [*y*] y [*z*] del grupo del nudo. Estas clases se llaman generadores del grupo.

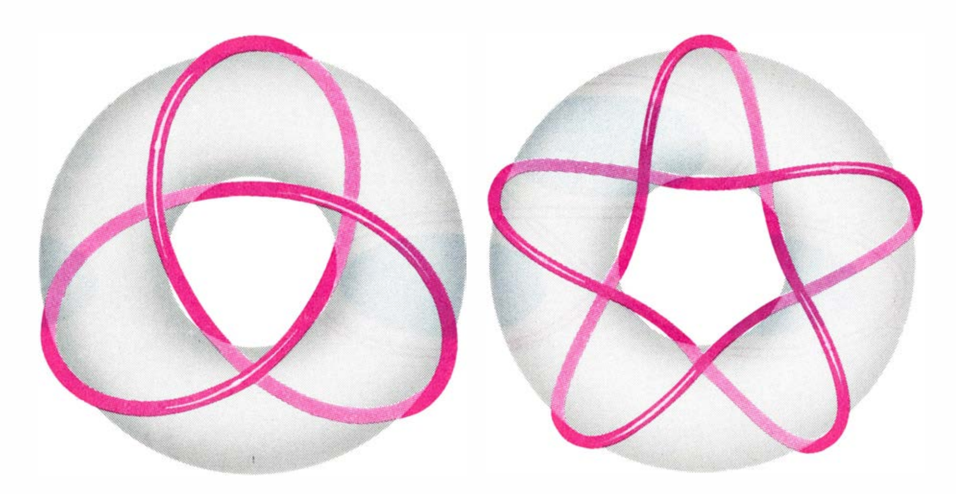
Observemos que en cada punto de cruce se reúnen los tres segmentos del nudo; uno de ellos por encima, y los otros dos, encontrándose bajo él. Por ejemplo, en el cruce de la parte superior izquierda, el segmento asociado al generador [*y*] se encuentra situado por encima, y los segmentos asociados a [*x*] y a [*z*] empalman bajo aquél. Así pues, en cada punto de cruce los generadores se conducen de distinta forma, lo que determina ciertas relaciones entre ellos. Podemos ver la forma concreta de la relación que liga los generadores en este cruce particular moviendo los extremos del camino *x* a lo largo del camino *y*, y como se muestra en la parte superior derecha de la ilustración. Se crea de esta forma un camino homotópico a *z* y también homotópico a $y^{-1}xy$. Así pues, la ecuación $[y]^{-1}[x][y] = [z]$, o como se expresa en la parte inferior derecha, $[y]^{-1}[z][x][y] = [e]$ manifiesta de qué forma se reúnen los generadores en el punto de cruce. El procedimiento se puede repetir en los otros dos puntos de cruce, obteniéndose las ecuaciones $[z]^{-1}[y]^{-1}[z][x] = [e]$ y $[x]^{-1}[z][x][y]^{-1} = [e]$, como se muestra en la ilustración de la página precedente. Tales ecuaciones se llaman relaciones del grupo nodal. En 1910, el matemático Max Dehn demostró que, dado un nudo cualquiera, los generadores y las relaciones obtenidas por este procedimiento dan una descripción completa del grupo nodal, en el sentido de que todo elemento del grupo es expresable como producto de los generadores y sus inversos, en una cierta forma, y que toda ecuación entre los elementos del grupo puede ser deducida a partir de las relaciones. Toda lista de generadores y relaciones que satisfaga estos requisitos se llama representación del grupo.

La transformación de un problema geométrico, concerniente al manejo de nudos en el espacio, en un problema puramente algebraico, relativo al manejo de ciertos símbolos, está terminada ahora. El método descrito para construir una presentación del grupo nodal de la hoja de trébol puede generalizarse para tratar nudos más complicados; existe cierto número de otros algoritmos para calcular explícitamente representaciones de grupos. Así pues, cualquiera que sea el nudo a estudiar, existe una presentación calculable de su correspondiente grupo y las relaciones de la representación pueden tratarse casi exactamente igual que las ecuaciones algebraicas ordinarias con varias incógnitas. La única diferencia reside en la necesidad de respetar el orden de las clases de homotopía que figuren en las relaciones. Pero aunque los elementos del grupo no conmuten, ambos miembros de una relación pueden ser multiplicados por unos mismos símbolos, y cada símbolo simplificarse con su inverso siempre que sean contiguos. (Es preciso notar que al introducir en una igualdad un símbolo nuevo debe hacerse por el mismo lado en cada miembro. Dicho de otra forma, si $[x] = [y]$, entonces $[a][x] = [a][y]$, pero en general no suele ser cierto que $[a][x] = [y][a]$. Análogamente, debido a que los caminos de las clases de homotopía de un grupo tienen sentido de recorrido, al tomar la inversa de un producto es necesario invertir el orden de los factores: $[xy]^{-1} = [y]^{-1}[x]^{-1}$.)

Así pues, en las presentaciones de un grupo pueden introducirse nuevos símbolos o eliminarse símbolos superfluos, resultando de aquí que un mismo y único grupo admite presentaciones de muy diverso aspecto. Por ejemplo, en la presentación del grupo para la hoja de trébol, [*z*] es expresable como producto de [*x*], [*y*] y sus inversos; por tanto, por simple sustitución y quizás alguna manipulación de las ecuaciones, es posible

eliminar $[z]$, obteniéndose una nueva presentación, perfectamente válida: $[x]$, $[y]$; $[x][y][x] = [y][x][y]$. (Por este procedimiento, al menos uno de los generadores de la presentación de Dehn para los grupos nodales podrá ser siempre eliminado.)

La presentación del grupo de la hoja de trébol puede transformarse todavía más, haciendo $[x][y]$ igual a $[u]$ y llamando $[v]$ al producto $[x][y][x]$. No es difícil demostrar entonces que cualquier producto de $[x]$, $[y]$ y sus inversos es expresable como producto de $[u]$ y $[v]$ y sus inversos, y que la relación $[x][y][x] = [y][x][y]$ puede ser transformada en $[u]^3 = [v]^2$. Así pues, el grupo nodal de la hoja de trébol puede describirse en todavía otra presentación más, que revela una sutil e insospechada relación entre los elementos del grupo: $[u]$, $[v]$; $[u]^3 = [v]^2$.



NUDOS TORICOS, que forman una clase de nudos cuyo grupo resulta ser invariante especialmente eficaz. Tales nudos están restringidos a la superficie de un toro (que se considera inmerso en el espacio tridimensional de la manera más natural), y tienen por ello grupos extraordinariamente sencillos y regulares, que los caracterizan casi totalmente. Los únicos nudos que el grupo nodal de un nudo tórico no es capaz de distinguir son el original y su simétrico respecto a un plano (que se obtiene cambiando todos los cruces por debajo del nudo primitivo por cruces por arriba, y recíprocamente) (*véase abajo*).

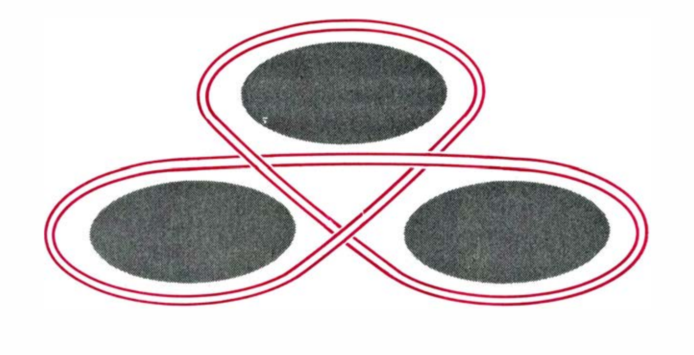
Aplicaciones del grupo nodal

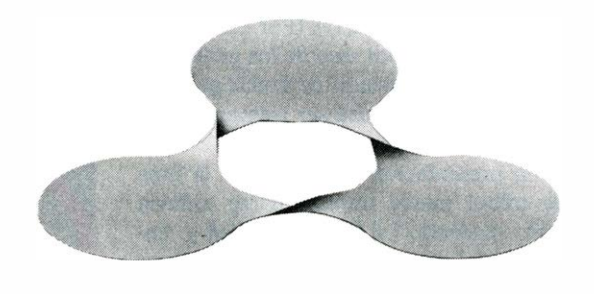
Dado que el objetivo de la teoría de nudos es diferenciar unos nudos de otros, el descubrimiento de que un mismo grupo nodal puede tener tan distintas presentaciones no deja de resultar decepcionante. Podríamos muy bien preguntarnos qué ventaja se ha obtenido de sustituir nudos por sus grupos nodales, o con mayor generalidad, de sustituir la geometría por álgebra. La ventaja es incuestionablemente importante. Para empezar, el grupo nodal resulta ser un invariante muy potente, capaz de distinguir muchos más nudos que, pongamos por caso, el mínimo número de puntos de cruce. No todos los nudos pueden distinguirse por sus grupos, pero solamente hace falta un poco más de información para determinar si dos nudos son o no equivalentes. Dicho de otra forma, los grupos nodales, para los que siempre es posible calcular una presentación, son casi lo suficientemente fuertes como para caracterizar los nudos.

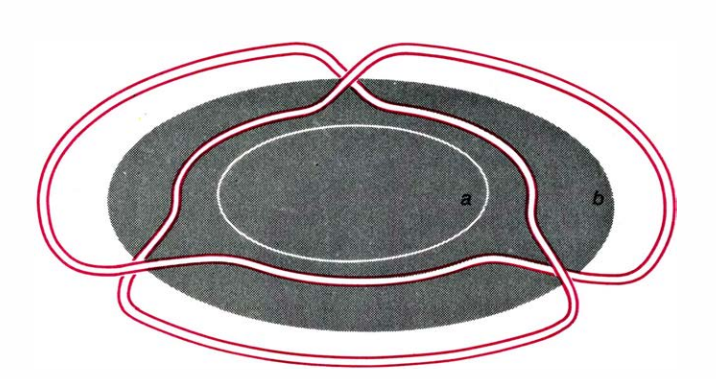
Además, el que las presentaciones de grupos idénticos puedan adoptar tan variados ropajes no es en realidad tan gran problema como podría parecer a primera vista. El problema de demostrar que dos objetos algebraicos son equivalentes (por ejemplo, confrontando sus miembros) puede resolverse frecuentemente por cálculo directo; hay además todo un amplio repertorio de teoremas puramente algebraicos que podrían aplicarse a tal fin. Por otra parte, quizá sea posible distinguir los grupos reconociendo propiedades características de ambos, en las que los grupos podrían diferir. Gran parte del aspecto algebraico de la teoría de nudos está consagrada a resolver este tipo de problemas.

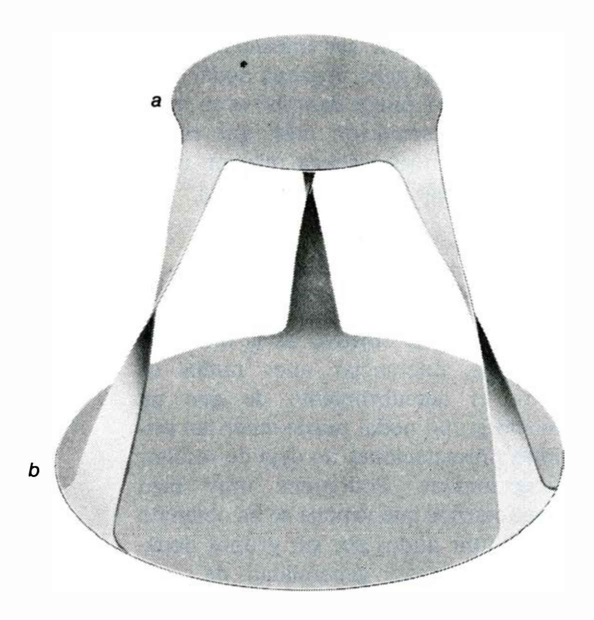


INTRODUCIENDO EL NUDO EN EL INTERIOR DEL TORO (*abajo, izquierda*) se ve que el nudo da tres vueltas según la circunferencia interior, señalada por el camino *a*. Encogiéndolo por el exterior (*abajo, derecha*) queda de manifiesto que da cinco vueltas alrededor de la circunferencia exterior, señalada por el camino *b*. Con toda generalidad, un nudo tórico cualquiera dará *p* vueltas según la circunferencia interior y *q* vueltas según la exterior. Por ejemplo, en la ilustración superior de esta página, el nudo de la izquierda tiene *p* igual a 2 y *q* igual a 3, mientras que en el de la derecha *p* es igual a 2 y *q* es igual a 5. La estereotipada presentación de todos los grupos de nudos tóricos obedece a esta propiedad característica. Obsérvese que los caminos *a* y *b* se encuentran en el complemento del nudo y tienen punto base común situado sobre el toro. El grupo de cualquier nudo tórico puede ser presentado mediante las clases de homotopía $[a]$ y $[b]$, que están relacionadas por una única y sencilla ecuación, $[a]^p = [b]^q$. Esta ecuación significa que en el complemento de un nudo tórico todo camino homotópico al *a*, bobinado *p* veces por dentro del toro, puede deformarse en un camino homotópico *b* bobinado *q* veces por el exterior del toro. No es difícil demostrar que los grupos de dos nudos tóricos son idénticos si y solamente si ambos nudos tienen el mismo par de valores de *p* y *q*. Así pues, el grupo nodal puede servir para determinar si dos nudos tóricos son equivalentes (salvo simetría respecto a un plano).

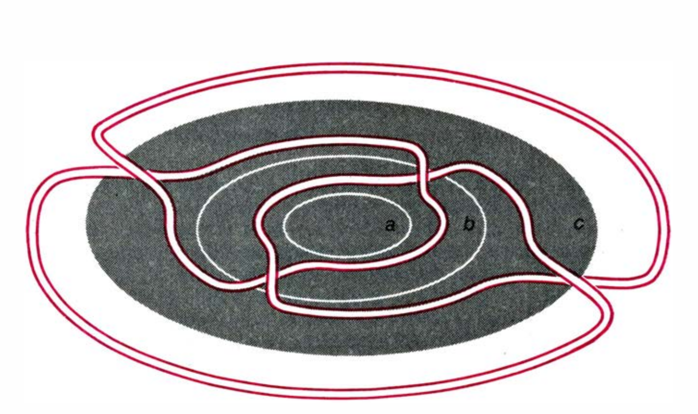


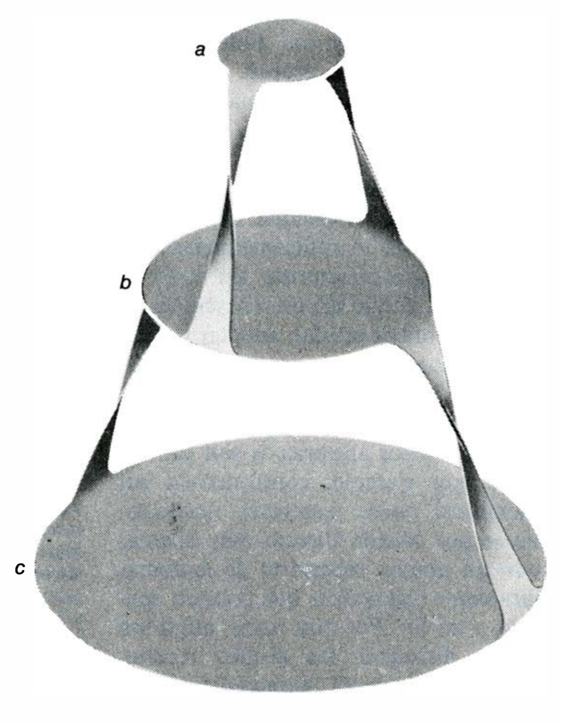






GEOMETRICAMENTE, UN NUDO puede considerarse como el borde único, o frontera, de una superficie bidimensional. Mostramos aquí dos superficies limitadas por el nudo en hoja de trébol. Ambas han sido construidas usando discos para "rellenar" los lazos del nudo, y cintas retorcidas media vuelta para rellenar las zonas de los puntos de cruce y conectar los discos. La construcción inferior se denomina "orientable", o de dos caras, porque al viajar de un nivel a otro se permanece en el mismo lado de la superficie (no ocurre así en la construcción superior, que es no orientable).



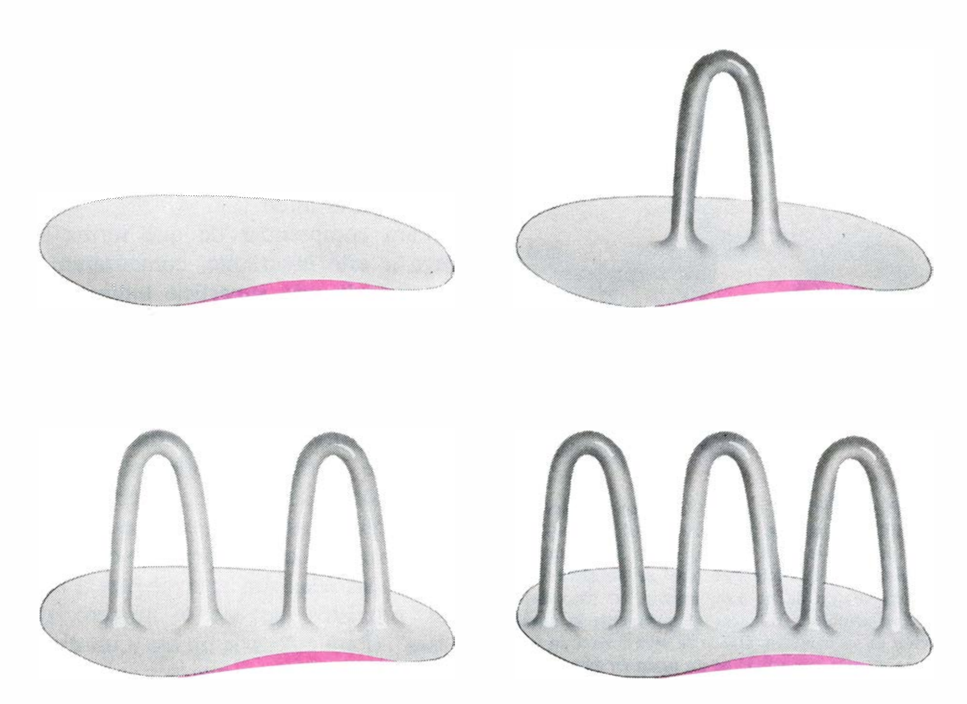


SUPERFICIE ORIENTABLE, que puede ser construida para nudos cualesquiera, de forma que el nudo sea el único borde de la superficie. El primer método general para construir tal superficie orientable, que mostramos aquí, fue ideado y puesto a punto por el matemático alemán H. Seifert, en 1935. La importancia de la construcción de Seifert, que dio origen a la definición de otro invariante del nudo, llamado género, reside en que toda superficie orientable es topológicamente equivalente a un disco provisto de un cierto número de asas (*véase la ilustración de la página siguiente*).

La fuerza del invariante "grupo nodal" se debe a que, y en contraste con el invariante "número mínimo de cruces", es fácil de calcular, y por otra parte, consigue en gran medida retener la complejidad y la estructura del nudo. Por esta razón, el invariante puede servir para contestar diversidad de cuestiones básicas o de carácter práctico acerca de los nudos, algunas de las cuales ya he mencionado más arriba. Por ejemplo, ¿es posible desanudar una curva anudada, es decir, hacerla equivalente a una circunferencia, formando en la curva un nudo más y desenmarañando después el resultado por algún método? Para responder a esta cuestión fijémonos en el grupo nodal asociado a la primera curva anudada. Al formar un nudo más en la curva, se originará un grupo nodal mayor y más complicado que el inicial. Por otra parte, como puede verse en la ilustración inferior de la página 62, el nudo trivial, o nudo desanudado, tiene el grupo más pequeño y sencillo: un grupo presentado por un único generador y carente de relaciones. Por consiguiente, el grupo resultante de añadir un nuevo nudo no podrá ser igual al grupo del nudo trivial, no siendo, pues, posible que un segundo nudo elimine al primero.

La observación de que el grupo del nudo trivial admite una representación tan poco complicada suscita una cuestión de lo más natural y engañosamente simple. ¿Está desanudado todo nudo que tenga esta misma representación mínima? Fueron precisos muchos años para resolver las profundas consecuencias implicadas por tan natural pregunta. En uno de los primeros artículos sobre teoría de nudos, publicado hacia 1910, Dehn mantenía haber demostrado que todo nudo cuyo grupo admita una presentación con un único generador es un nudo trivial; pero en 1928, el matemático alemán H. Kneser descubrió que la demostración de un resultado intermedio utilizado por Dehn para probar su teorema era insuficiente. Pasaron otros 29 años hasta que C. D. Papakyriakopoulos, de la Universidad de Princeton, pudo demostrar correctamente aquel resultado (que, irónicamente, ha venido a conocerse como lema de Dehn), y demostrar por tanto el teorema.

El teorema de Dehn es un resultado útil e importante. Sin embargo, no resuelve el problema de decidir si un nudo arbitrario está anudado o no, pues no siempre es fácil determinar si un grupo nodal es equivalente al grupo del nudo trivial. Lo que sí establece el teorema de Dehn es que el nudo trivial se distingue



PARA DOTAR DE ASAS A UN DISCO se perforan en él dos orificios y se conecta un tubo de uno a otro. En estos ejemplos las asas se disponen tan sencillamente que los bordes de los discos no se anudan. Las asas de superficies limitadas por curvas anudadas se enlazarían en forma más compleja.

por quedar totalmente caracterizado por su grupo. Hay otra clase de nudos para los que casi puede mantenerse este mismo aserto. Los nudos de esta clase son los que pueden dibujarse sobre la superficie de un toro. Un toro es una figura geométrica, semejante a un neumático inflado, que se engendra haciendo girar una circunferencia alrededor de una recta contenida en su mismo plano, y exterior a ella.

Nudos tóricos

Los nudos tóricos se presentan al estudiar ecuaciones con dos variables complejas. (La rama de la matemática que estudia las formas geométricas originadas por consideraciones algebraicas se llama geometría algebraica.) Debido a que estos nudos están restringidos a la superficie del toro, que se considera inmersa en el espacio tridimensional en la forma más natural, sus grupos nodales tienen presentaciones sencillas y estereotipadas. Además, los nudos tóricos quedan casi totalmente caracterizados por sus grupos nodales.

En la ilustración superior de la página 63, se dan algunos ejemplos de nudos tóricos. Como se muestra en la ilustración inferior de esa misma página, cuando se traza sobre una superficie tórica uno de estos nudos, da p vueltas alrededor de la circunferencia interior del toro, y q vueltas alrededor de la circunferencia exterior, siendo p y q enteros positivos. Esta fundamental característica de todos los nudos tóricos es la causa última de la extraordinaria regularidad de sus presentaciones. A partir de un punto base fijo, situado sobre la superficie del toro y en el complemento del nudo, se trazan caminos que rodeen al toro. El primero da la vuelta al toro por su interior y según la circunferencia más larga; el segundo, lo rodea por el exterior, siguiendo la circunferencia más corta. Si estos caminos se denotan a y b, como muestra la ilustración, las clases de homotopía $[a]$ y $[b]$ sirven de generadores del grupo nodal tórico. Estos generadores están ligados por la sencilla relación $[a]^p = [b]^q$. Dicho de otra forma, en el complemento del nudo todo camino homotópico al camino formado por una bobina de p espiras idénticas a a trazadas por el interior del toro puede ser deformado hasta dar un camino homotópico a una bobina de q espiras idénticas a b que rodee exteriormente al toro. (Unos cuantos garabatos servirán para hacernos caer en que p y q han de ser primos entre sí, es decir, que es imposible trazar un nudo sobre el toro donde p y q admitan algún divisor común.)

Los nudos tóricos son los únicos nudos cuyos grupos están presentados por sólo dos generadores más esta única y sencilla relación definidora. Por otra parte, no es difícil demostrar que dos grupos tales únicamente serán equivalentes cuando tengan igual par de valores de p y q. Así pues, los grupos tóricos pueden ser clasificados completa y eficazmente de acuerdo con su grupo. Es preciso notar, sin embargo, que el grupo



GENERO DE UN NUDO es, por definición, el mínimo número de asas que exigiría una superficie orientable para que su único borde fuese el nudo. La superficie que aquí se muestra es un disco dotado de un asa, aunque ha sido deformada al objeto de hacer más evidente la anudación de su frontera. Dicho de otra forma, la frontera de esta superficie es un nudo de género igual a 1 o menor. Siendo imposible situar este nudo particular sobre el contorno de un disco, su género es realmente 1. El género es un invariante del nudo, y aunque no tan potente como el invariante "grupo nodal", ha resultado útil para elucidar importantes cuestiones de teoría de nudos.

no es capaz de distinguir dos nudos tóricos simétricos respecto a un plano, que se deducen uno de otro cambiando cada cruce por debajo en un cruce por encima, y viceversa.

Los grupos de nudos tóricos son probablemente los menos complicados de todos los grupos nodales, a excepción del grupo del nudo trivial. Quizás, en consecuencia, tales grupos parezcan tener más organización que la mayoría de grupos nodales. Por ejemplo, los grupos tóricos son los únicos grupos nodales no triviales donde existen elementos distintos de la identidad que conmutan con todos los restantes elementos del grupo. Con otras palabras, existe en ellos al menos un elemento no trivial $[c]$ tal que $[c]\,[x] = [x]\,[c]$ para toda clase $[x]$ de homotopía del grupo nodal. (En vista de que las potencias de $[a]$ conmutan con $[a]$ y de que las potencias de $[b]$ conmutan con $[b]$, fijándonos en la relación $[a]^p = [b]^q$ no es difícil reconocer un tal elemento $[c]$: es precisamente $[a]^p$. Así pues, entre los grupos nodales no triviales, los grupos de nudos tóricos son los que más cerca están de mostrar conmutatividad completa.

Un invariante geométrico

Los nudos tóricos demuestran de forma impresionante hasta qué punto puede resultar efectivo un invariante algebraico; pero existen también invariantes de otros tipos que esclarecen los nudos y su teoría. Concretamente, se ha utilizado un invariante geométrico para solucionar ciertas cuestiones teóricas relativas a nudos. Tal invariante se llama género del nudo, y para definirlo se construye una superficie cuyo único borde sea el nudo.

Para comprender de qué forma se llega a este invariante, comenzaremos imaginando una superficie bidimensional de un solo borde, inmersa en el espacio tridimensional. Las superficies de este tipo pueden tener, bien sea dos caras, o bien sólo una cara, como la banda de Möbius. Las superficies de dos caras se llaman también superficies orientables, y las de una, superficies no orientables. Las superficies orientables son de manejo particularmente cómodo, pues todas ellas son equivalentes a discos provistos de cierto número de "asas". (Para colocarle un asa a un disco se cortan en él dos agujeros y se conecta un tubo desde un orificio al otro.) Retornemos ahora a la construcción de una superficie cuyo borde sea un nudo dado. La ilustración superior de la página 64 pone de manifiesto que tal superficie puede ser orientable o no serlo. No obstante, el matemático alemán H. Seifert ideó en 1935 un procedimiento para construir una superficie orientable cuyo único borde sea un nudo dado, válido cualquiera que sea el nudo.

La construcción de Seifert probaba que todo nudo ha de ser borde de al menos una superficie equivalente a un disco provisto de asas. En realidad, dado un nudo cualquiera, habrá muchas superficies de este tipo, dotadas de distintos números de asas. Entre todas ellas, la que menos asas tenga se llamará superficie minimal del nudo. El género se define como el número de asas de una superficie minimal. Es imposible disminuir el número de asas deformando el nudo y, por tanto, el género del nudo es verdadero invariante del nudo.

El género, y la superficie minimal asociada, han demostrado ser valiosos instrumentos en el análisis de diversos problemas de la teoría. Por ejemplo, si bien el género no es un invariante tan potente como el grupo nodal, sí ha demostrado mayor eficacia para determinar si un nudo dado está anudado o no. Es evidente que el nudo trivial o desanudado –la circunferencia– es borde de un disco sin asas, y por tanto es de género cero. Además, puede demostrarse que el nudo trivial es el único con este género. Resulta así posible demostrar el auténtico carácter de "anudación" para todo nudo cuyo género sea calculable. No obstante, por lo común el género es difícil de calcular, por lo que no parece presentar ventaja sobre el grupo nodal.

En 1962, no obstante, Wolfgang Haken, de la Universidad de Illinois en Urbana-Champagne, ideó un algoritmo con el fin de calcular el género de un nudo dado a partir de un dibujo. Además, el año pasado Friedhelm Waldhausen, de la Universidad de Bielefeld, se sirvió de resultados de Haken y G. Hemion, de Bielefeld, para construir un algoritmo cuyo fin es determinar si dos dibujos representan o no nudos equivalentes. La verdad es que estos algoritmos resultan, con mucho, enormes para poder ser utilizados, con lo que, en la práctica, no existe ningún procedimiento para distinguir los nudos verdaderos de los triviales, o para diferenciar unos nudos de otros. Sin embargo, los anteriores algoritmos muestran que estos problemas son decidibles: se podría, al menos en principio, construir una máquina que aceptase el dibujo de un nudo cualquiera, calculase su género y decidiese si el nudo estaba anudado o no. O bien, se podría construir una máquina que, dados los dibujos de dos nudos, decidiría si éstos eran o no equivalentes.

El género de un nudo es un potente invariante geométrico, lo mismo que el grupo nodal es un potente invariante algebraico. Uno de los más interesantes descubrimientos recientes de la teoría de nudos ha sido la conexión existente entre ambas construcciones matemáticas. En efecto, Charles D. Feustel, del Virginia Polytechnical Institute, y Wilbur Whitten, de la Universidad de Southwestern Louisiana, han demostrado que el género, cuya definición es puramente geométrica, es también un invariante algebraico determinado por el grupo nodal. Una vez más, la interacción de geometría y álgebra da a la teoría de nudos profundidad e interés.

He intentado en este artículo mostrar cómo se han resuelto algunos problemas fundamentales de la teoría de nudos, y cómo se están atacando algunos otros. En la actualidad, sin embargo, para el estudio de los nudos los matemáticos están tomando muchas otras direcciones. Por ejemplo, en este momento se están empleando técnicas de teoría de nudos para construir y estudiar diversidad de interesantes espacios tridimensionales. Además, el problema de los nudos en dimensión superior –superficies bidimensionales anudadas en espacios de dimensión cuatro, variedades tridimensionales en espacios de dimensión cinco, etcétera– está recibiendo mucha atención. Muchos resultados de la teoría de nudos clásica pueden generalizarse a dimensión superior, abriendo de esta forma nuevos, ricos y complejos temas de investigación matemática.

Una mina de sílex neolítica

El comienzo de la agricultura aumentó la demanda de útiles de sílex. Las excavaciones llevadas a cabo en Holanda en un valle rico en sílex han permitido conocer que los mineros de hace más de 5000 años lo extraían a toneladas

Peter W. Bosch

De los minerales silíceos, el sílex o pedernal ha constituido un hito importante, de al menos medio millón de años, en la historia de la humanidad. Más duro que el acero, es también quebradizo y por tanto fácil de trabajar por percusión o pulimento, con lo que se le puede dar cualquier forma que se desee. Sus nódulos, algunos de hasta un metro de diámetro, son muy abundantes en los lechos calcáreos depositados en los fondos oceánicos de todo el mundo durante el Cretácico Superior, hade unos 80 millones de años. En Europa se pueden encontrar lechos calcáreos del Cretácico al descubierto en diversos lugares que van en dirección este desde Francia y el sur de Inglaterra hacia Polonia y Checoslovaquia.

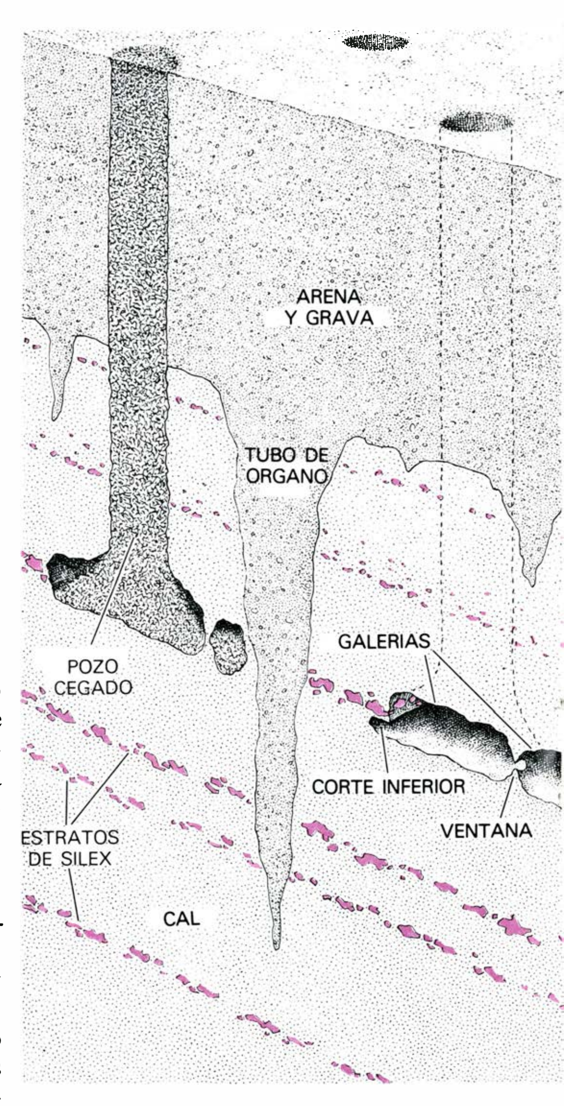
Durante el Paleolítico, los cazadores de Europa tenían preferencia por los nódulos sueltos de sílex, que se habían desprendido por causas naturales del lecho calizo, como materia prima para hacer hojas, raspadores y otros útiles. Si nos remontamos algunos miles de años atrás, con la aparición de la agricultura como modo de vida, la creciente demanda de instrumentos de sílex hizo que los europeos del Neolítico pasasen de recoger los nódulos de sílex de la superficie a una explotación minera. En este artículo expondré lo que se ha averiguado recientemente sobre uno de estos yacimientos mineros, de donde los obreros del Neolítico (hace unos 5000 años) extrajeron material para producir más de 150 millones de útiles. El yacimiento, cerca de Rijckholt, en Holanda, se descubrió en 1881. A partir de entonces se estudió de forma irregular durante setenta y cinco años. Por último, un grupo de arqueólogos aficionados lo excavaron exhaustivamente trabajando una noche por semana a lo largo de nueve años.

El río Mosa nace en Francia y desemboca en el Mar del Norte, después de atravesar Bélgica y el ángulo suroeste de Holanda cerca de Maastricht. Marcel de Puydt, prestigioso arqueólogo belga, se dio cuenta en un viaje que realizaba por tren desde la ciudad fronteriza belga de Visé a Maastricht en 1881, de que la forma del terreno del valle del Mosa era muy parecida a otros lugares de su país donde había localizado yacimientos prehistóricos. El paisaje arbolado ascendía gradualmente desde el valle hasta terminar en una meseta deforestada. De Puydt decidió hacer un reconocimiento de la zona y ese mismo año encontró un depósito de instrumentos de sílex en superficie cerca de Rijckholt.

Por tanto, debe considerarse a De Puydt como el descubridor del yacimiento de Rijckholt, a pesar de que tardaría todavía algún tiempo en encontrar la zona principal de trabajo. Era ésta una hoya natural, llena de lascas de sílex desgastado por el uso, de una profundidad de metro y medio, que más tarde recibió el nombre de "Gran Taller". En cualquier caso, y aunque otros vinieron a investigar el lugar, De Puydt continuó trabajando en Rijckholt esporádicamente a lo largo de 35 años, ampliando su colección de útiles de sílex (hachas, raspadores, puntas de proyectiles y hojas de cuchillos) e instrumentos de minería (picos hechos de asta de ciervo).

René de Geloes, propietario del terreno donde estaban situadas las minas de Rijckholt, también investigó el yacimiento. El conde Geloes empezó a trabajar en 1886; entre sus colaboradores estaba el joven estudiante de medicina Eugène Dubois, que cuatro años después descubrió en Java el *Pithecanthropus erectus* (denominado *Homo erectus* en la actualidad). También trabajaron en Rijckholt otros arqueólogos belgas y franceses, como Joseph Hamal Nandrin, de la Universidad de Lieja, que comenzó a excavar en 1903 con la ayuda de los aldeanos locales. En 1910 Nandrin encontró varios pozos de mina antiguos llenos de derrubios calcáreos. Este descubrimiento proporcionó la primera prueba consistente de que habían existido minas prehistóricas de sílex en Holanda (aunque el Gran Taller de Rijckholt, con su enorme cantidad de desechos de sílex, constituía una razón vigorosa de que se habían explotado en las cercanías nódulos de sílex).

Hasta 1923 ningún investigador ho-



CORTE ESQUEMATICO de la región minera de sílex de Rijckholt, en las proximidades de Maastricht, en el extremo meridional de Holanda. A la izquierda, capa de arena y grava que cubre un

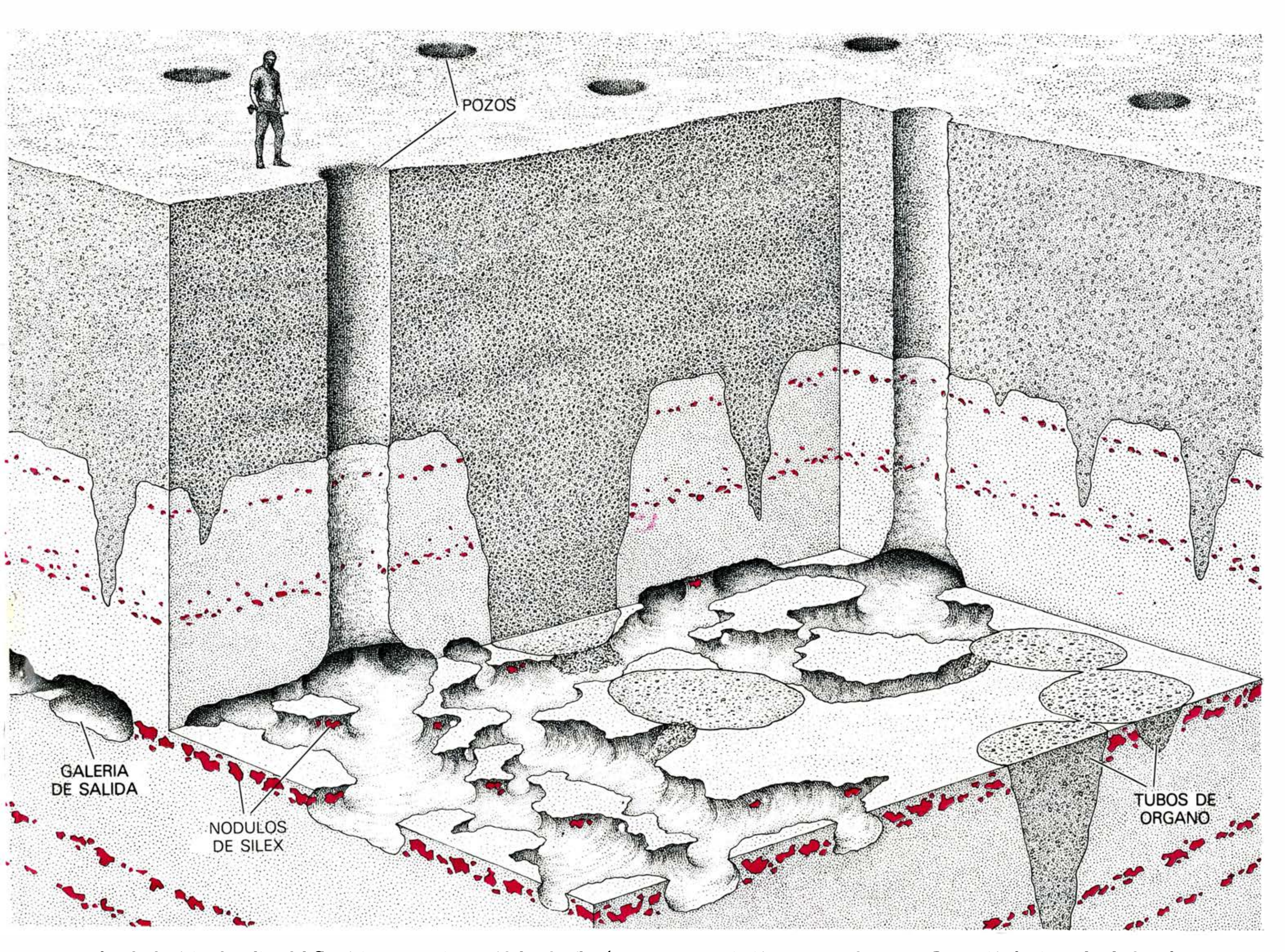
landés había demostrado interés por las minas de sílex de Rijckholt. Pero en ese mismo año A. E. van Giffen, de la Universidad de Groningen, inició un estudio sistemático del yacimiento. Fue el primero en trazar una serie de zanjas en el Gran Taller y determinó cuidadosamente el emplazamiento de los hallazgos. Las zanjas demostraban que los neolíticos habían dividido la explotación en una serie de zonas separadas. Por ejemplo, las hachas se hacían en unos lugares determinados, los cuchillos en otros, y los raspadores en otros; cada zona de trabajo especializado podía identificarse por el tipo de láminas y lascas característicos de un proceso de producción. Van Giffen también estudió el yacimiento de Rijckholt en su conjunto. Allí donde el perfil de la meseta descendía hacia un extremo del Gran Taller, encontró rastros de pozos mineros. En la misma meseta había restos de cerámica, huellas de hogares y lo que parecían ser señales de posibles cobertizos. Ello venía a indicar que los mineros habían acampado allí mientras excavaban el sílex.

Van Giffen terminó el trabajo de campo en 1926. A sugerencia de algunos arqueólogos belgas, los frailes dominicos del convento de Rijckholt continuaron investigando el lugar en su tiempo libre. Entre 1929 y 1932 hallaron también restos de pozos mineros en la pendiente cerca del Gran Taller. Descubrieron además restos de galerías hundidas (túneles horizontales que se excavan hacia afuera desde el fondo de un pozo) y recogieron cierto número de útiles de minería: más de 100 picos de asta de ciervo y unas 1200 hachas de sílex. Durante este período, Nandrin solía presentarse desde Lieja para recoger instrumentos de Rijckholt con destino a su universidad, lo que continuó haciendo hasta 1953.

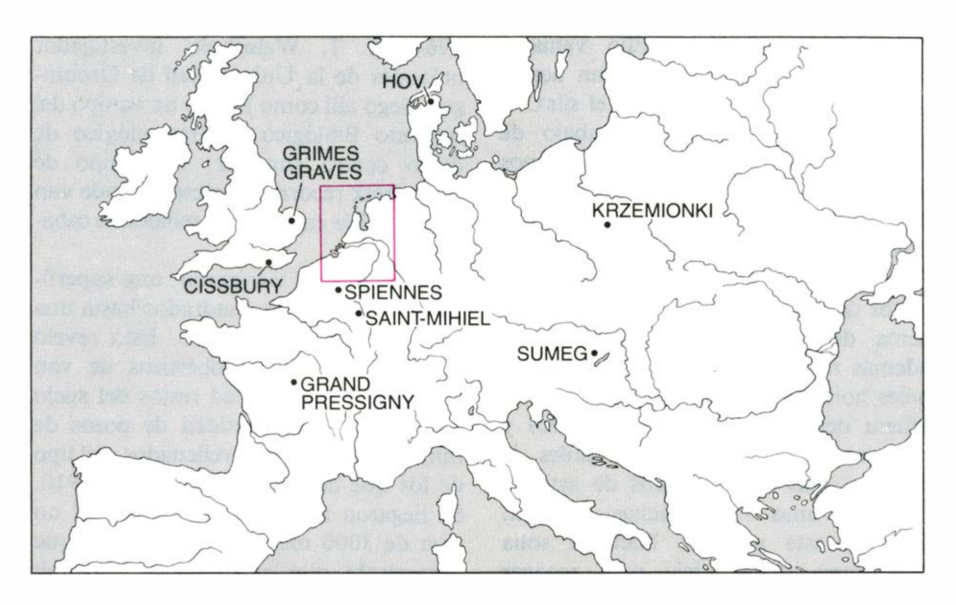
Los trabajos en Rijckholt se interrumpieron durante la década siguiente. En 1964, H. T. Waterbolk, investigador holandés de la Universidad de Groningen, llegó allí como jefe de un equipo del Instituto Biológico y Arqueológico de dicho centro superior. El equipo de Waterbolk recorrió la meseta donde van Giffen había encontrado señales de cabañas.

Limpiaron, y tamizaron, una superficie de 4000 metros cuadrados hasta una profundidad de un metro. Esto reveló que las huellas de cobertizos de van Giffen eran en realidad restos del suelo que cubría gran cantidad de pozos de mina que habían sido rellenados, del tipo de los que descubrió Nandrin en 1910. Se llegaron a localizar 66 pozos en un área de 3000 metros cuadrados, lo que demostraba que las minas de Rijckholt habían ocupado una región mucho más extensa de lo que habían supuesto los primeros investigadores.

Tanto Waterbolk y sus colaboradores como antes van Giffen y los frailes



hondo depósito de caliza del Cretácico que encierra nódulos de sílex (*en color*). Los "tubos de órganos" son zonas donde la infiltración de agua disuelve la caliza, ocupando su lugar intrusivamente la capa de arena y grava. La sección vertical revela la existencia de pozos de una sola boca y las galerías su asociación con otros dos pozos. La sección horizontal, a la derecha, muestra dos pozos y la mitad inferior de las galerías que están asociadas con ellos, según el trazado de los arqueólogos aficionados que volvieron a abrir las minas neolíticas en 1964. La figura humana sirve como escala de referencia.



PRINCIPALES MINAS de sílex de la Europa neolítica. Entre ellas, se encuentran las de Grimes Graves y Cissbury en Inglaterra, Grand Pressigny y Saint-Mihiel en Francia, Spiennes en Bélgica, Hov en Dinamarca, Krzemionki en Polonia y Sumeg en Checoslovaquia. (*En color, la ilustración inferior.*)



MINAS HOLANDESAS DE SILEX, ubicadas entre Maastricht y Aquisgrán, en las proximidades de un extenso depósito calcáreo (*en color*). Se debe al arqueólogo belga Marcel de Puydt el descubrimiento, en 1881, de que el hombre prehistórico del Neolítico fabricó ingentes cantidades de útiles de sílex.

dominicos comprobaron la existencia de galerías hundidas a lo largo de la pendiente junto al Gran Taller. Trataron de abrir un pasadizo para entrar en los lugares de trabajo neolíticos, pero no lo consiguieron porque ninguno de ellos tenía experiencia práctica en minería. Las investigaciones se interrumpieron, y parecía que nadie habría de llegar nunca a saber cómo habían sido en realidad las minas de Rijckholt.

Así las cosas, un geólogo profesional y entusiasta aficionado a la arqueología vino a hacer una nueva propuesta a Waterbolk. Se trataba de Werner M. Felder, del Instituto de Investigación Geológica de Holanda, miembro activo de la sección de Limburgo de la Sociedad Geológica Holandesa. Propuso abrir un túnel en la pendiente por encima del Gran Taller para seguir el estrato calcáreo rico en sílex que los mineros neolíticos habían explotado abriendo pozos en la meseta. Dicho túnel cortaría seguramente muchas de las galerías horizontales que los mineros habían excavado a través de la creta para extraer los nódulos de sílex. Felder calculó que bastaba con que el túnel de inspección tuviese unos 140 metros de largo. Propuso explorar también los 10 primeros metros de cada galería prehistórica cortada por el túnel.

Waterbolk aprobó el proyecto de Felder y se ofreció como supervisor científico. Las minas de Rijckholt, que no pertenecen ya a los herederos del Conde de Geloes, están situadas en una reserva natural local. El administrador autorizó una excavación testigo, y 13 miembros de la Sociedad Geológica (sección de Limburgo) formaron un equipo al que denominaron "Grupo de trabajo sobre minería de sílex prehistórica". Estaba integrado por geólogos, obreros, estudiantes, un profesor y, por supuesto, varios mineros veteranos de las minas de carbón cercanas. Aunque la procedencia de los miembros del grupo era tan variopinta, a todos les unía tres intereses: la arqueología, la minería y el deseo de hacer ellos mismos la excavación. Y así fue como nació la compañía minera más pequeña y extraoficial de Holanda.

Cuando empezamos a trabajar el 6 de junio de 1964, nadie tenía idea del volumen de la tarea, cuánto tiempo duraría o cuánto costaría. Por la misma razón, nadie sabía qué aspecto tendrían las minas neolíticas cuando llegáramos a ellas. Todos creíamos que las galerías prehistóricas serían lo bastante altas como para andar de pie por ellas; los optimistas del grupo venían incluso

equipados con el material necesario para cartografiar. Las cosas resultaron ser muy distintas.

Lo primero que hicimos fue construir un cobertizo en el punto de entrada sobre la cuesta. Esta precaución resultó útil porque hacía muy mal tiempo cuando nos pusimos manos a la obra con picos y palas para abrir una boca de túnel de unos dos metros de ancho por uno y medio de alto. Como nuestro trabajo era voluntario y gratuito, sólo podíamos dedicar a la excavación los fines de semana. Sin embargo, antes de que el túnel midiese unos pocos metros, encontramos las primeras galerías neolíticas, y al cabo de varios fines de semana llegamos al primer pozo.

El pozo y las galerías estaban llenos de derrubio calcáreo; además, éstas se hallaban parcialmente hundidas. A medida que extraíamos los derrubios, íbamos metiendo vigas para entibar el techo de las galerías y las entradas a los pozos. Las primeras hachas de sílex las encontramos mientras limpiábamos las galerías. Al sacar el relleno del primer pozo también encontramos bastante ceniza, procedente al parecer de leña que cayó de una hoguera prehistórica situada junto a la boca del pozo; enviamos las cenizas para que se determinase la antigüedad por el carbono 14.

Esperamos los resultados con impaciencia. Durante 80 años se venía investigando el yacimiento de Rijckholt y nadie sabía aún cuál era la antigüedad de las minas. La respuesta que nos enviaron las fechaba en el año 3150 a. C. (con un margen de error de 60 años). Sin duda alguna, la explotación minera de sílex en el sur de Holanda empezó hace algo más de 5000 años.

Con la excavación testigo nuestro túnel se adentró 15 metros en la creta. Los resultados eran lo bastante estimulantes para garantizar un proyecto de mayor envergadura, por lo que se necesitaba un permiso oficial del gobierno aunque el trabajo lo realizásemos aficionados. El permiso dependía de que utilizásemos entibados metálicos en vez de los de madera, y la Fundación Holandesa para Investigación Científica nos proporcionó una subvención que ayudó a hacer frente a estos gastos de sustitución.

Ahora que el proyecto había adquirido rango oficial, había que imponer una serie de reglas formales. Los grupos de trabajo de la excavación testigo venían al yacimiento los sábados, pero la mayoría estábamos casados y queríamos pasar el fin de semana con nuestras familias. Se decidió adelantar el horario de trabajo, y empezar el viernes a las 7 de la tarde hasta las 3 de la madrugada del sábado. También se acordó que cada miembro del grupo trabajase un mínimo de 20 noches al año. Más adelante, 20 voluntarios formaron el cuadro permanente de la sección minera de Limburgo, de manera que había equipos de 10 hombres trabajando de noche en Rijckholt durante 40 de las 52 semanas del año. Incluso si la nochebuena o nochevieja caían en viernes la mayoría de voluntarios venían a trabajar.

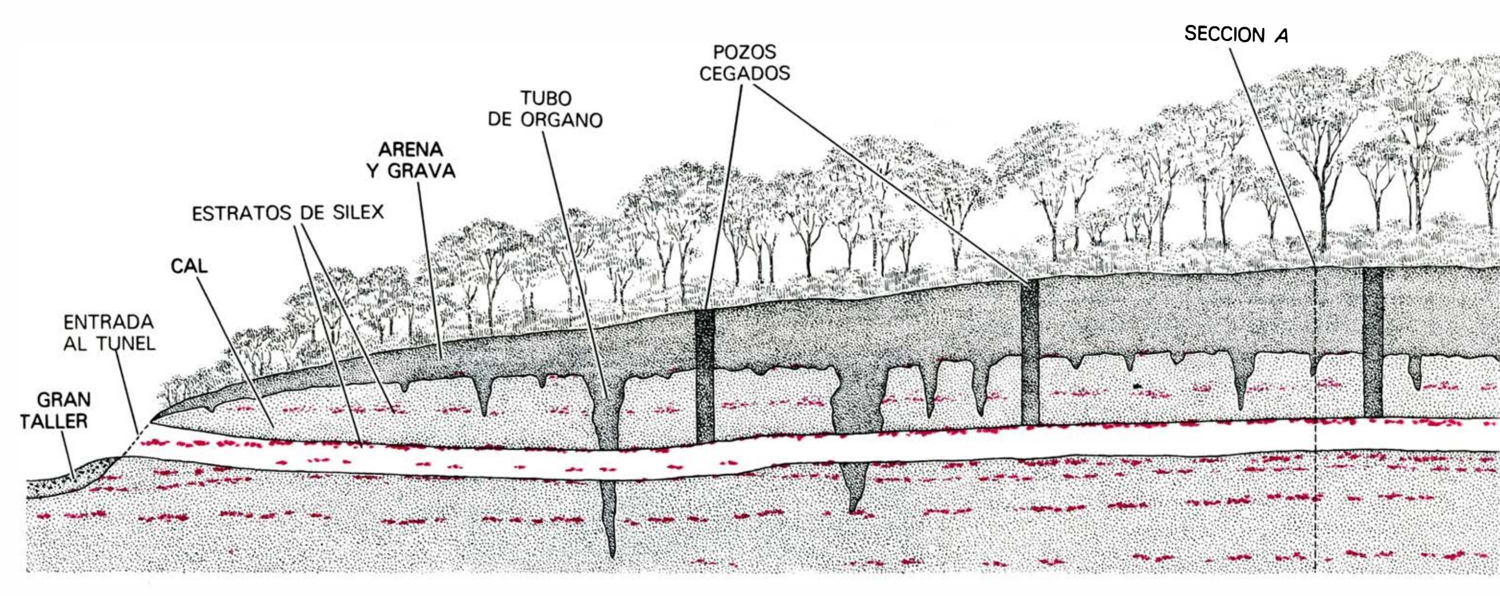
Werner Felder y su hermano Joseph fueron elegidos jefe y director técnico del proyecto, respectivamente. Los dos habían trabajado ya en minas de carbón holandesas y, por tanto, tenían experiencia minera práctica. Otros miembros del grupo se ocuparon de las relaciones con



GALERIA DE MINA en fase de limpieza de la cal de derrubio acumulada allí y que los mineros del Neolítico no terminaron de abrir. Mide unos 60 centímetros de altura. Las inclusiones negras en la pared caliza (*izquierda*) son nódulos de sílex. Estos abundan en los lechos calcáreos del Cretácico.



DOS "VENTANAS" abiertas en la pared de otra galería minera. Resultan demasiado pequeñas para servir de paso a otra galería. Quizá fueron fortuitas. La galería mide 80 centímetros de altura.



TUNEL DE EXPLORACION excavado por voluntarios de fin de semana entre junio de 1964 y diciembre de 1972. Se ilustra en elevación y sección transversal (*al final de la página opuesta*). La figura humana sirve de escala de referencia. Con una longitud de 150 metros a partir de la entrada al mismo, y adyacente aquella al Gran Taller, el túnel seguía la capa o nivel calizo de 30 centímetros que los mineros del Neolítico habían hallado parti-

instituciones y museos que se habían interesado por el proyecto.

En resumen, el proceso fue así:

En 1965, se acarrearon hasta el yacimiento 50 toneladas de acero para sustituir a los entibados de madera que se habían utilizado durante la excavación testigo.

Se instalaron unos raíles para extraer de la mina la masa caliza en vagonetas, que remplazaron a las carretillas. Para final de año, el túnel medía ya 18 metros de largo.

En 1966, una de las compañías interesadas en el proyecto nos prestó un compresor de aire y martillos neumáticos para facilitar el trabajo en la masa calcárea que era cada vez más dura. Durante el año anterior, nuestra única fuente de energía eléctrica había sido una batería de 12 volt; ahora nos prestaron un pequeño generador y pudimos montar un sistema de iluminación más extenso. A medida que el trabajo avanzaba, se hizo necesario dividir los grupos del viernes por la noche en equipos más reducidos; a imagen del sistema seguido en las minas de carbón, al frente de cada equipo había un jefe responsable de la seguridad del grupo que, además, llevaba el registro de los hallazgos. Para final de año, el túnel medía ya 38 metros de longitud.

En 1967, se hizo cada vez más difícil transportar la caliza sobrante. La industria holandesa de minas de carbón, que había mostrado gran interés por nuestro proyecto, nos prestó una correa transportadora de 150 metros de largo. Como no había energía eléctrica en el yacimiento, compramos un tractor y conectamos una rueda trasera con la correa transportadora mediante una polea. Gracias a las subvenciones económicas y a los equipos con que nos obsequiaron, pudimos hacernos con trajes de minero, cascos de seguridad y linternas. Además, a todos se nos hizo un seguro contra accidentes ya que el trabajo era arriesgado. Para final de año el túnel medía ya 67 metros.

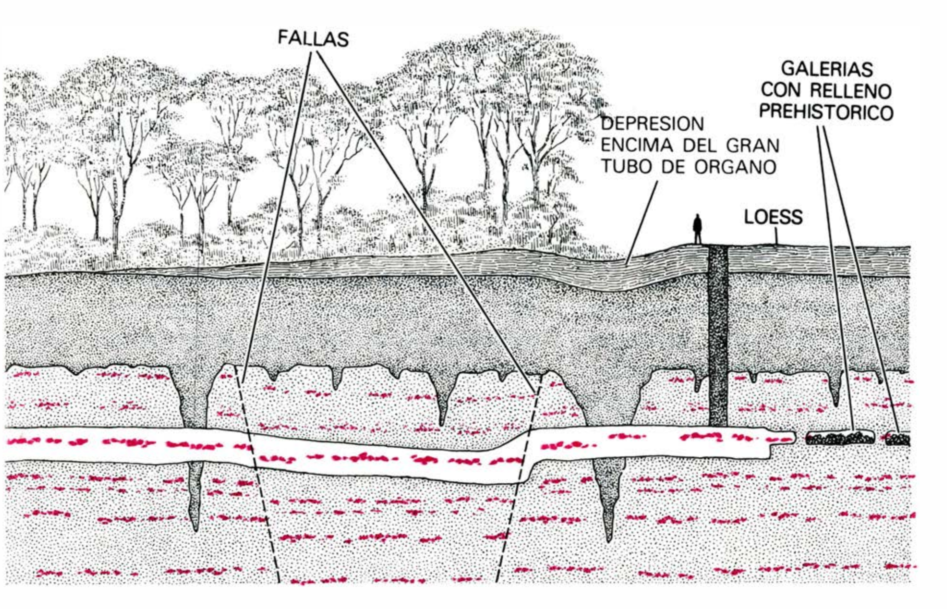
En 1968 continuamos trabajando en equipos. Uno de los cuales operaba en el túnel, otro se ocupaba del mantenimiento del material y la limpieza del yacimiento, y un tercero –y a veces un cuarto– desescombraba las galerías laterales, cartografiaba las zonas explotadas y llevaba el registro de los hallazgos y cenizas que se encontraban. Al acabar ese año, el túnel medía ya 108 metros de largo.

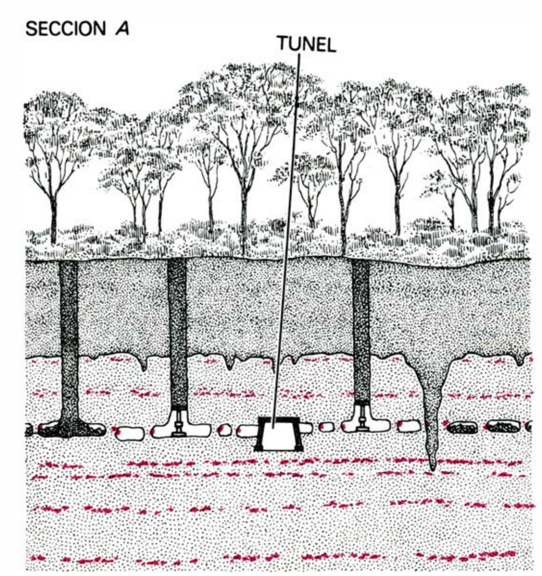
En 1969, de acuerdo con Waterbolk, se decidió alargar el túnel hasta cubrir toda la explotación minera neolítica. Esto suponía excavar 40 metros más y en todos reinaría el entusiasmo. Algunas de las galerías recién descubiertas se dejaban sin explorar ya que todo el trabajo se concentraba en profundizar en el túnel. Sin embargo, antes de fin de año ya se habían desescombrado las tres cuartas partes de todas las galerías hasta una distancia de 10 metros, y se habían recuperado unos 10.000 instrumentos de sílex. Para finales de ese año de 1969, el túnel perforado medía unos 110 metros de largo.

En 1970, 1971 y 1972 se continuó trabajando tanto en la excavación como en casa, lo que implicaba recolectar fondos (por entonces el gobierno había empezado a cerrar muchas de las minas de carbón que nos ayudaban), mantener correspondencia, topografiar el yacimiento y catalogar los hallazgos. Los pocos metros que faltaban por excavar nos plantearon un problema ya antiguo pero a otra dimensión.

Se llaman "tubos de órgano" a unas perforaciones formadas por la penetración del agua del suelo a través de la caliza; su diámetro varía desde un metro a más de 10; el espacio que antes ocupaba la caliza se rellena ahora de arena y grava. Si un excavador llega a tropezar con un tubo de órgano al hacer el túnel, existe el peligro de que el relleno caiga en el túnel bloqueando el paso o incluso enterrándole. Había muchos tubos de órgano por toda la zona de Rijckholt, pero afortunadamente los que encontramos eran pequeños. En las postrimerías de nuestro trabajo aparecieron algunos de dimensiones peligrosas. Sin embargo, antes de finalizar 1972 el túnel había alcanzado la longitud planeada de 150 metros.

En 1973, excavada ya toda la zona minera, pudimos dedicar más tiempo a limpiar las galerías restantes, lo que conseguimos tener hecho para final de año. Joseph Felder calculó que habíamos abierto un área minera de 3000 metros cuadrados. Lo único que nos quedaba por hacer era retirar la maquinaria y material de trabajo y reparar la superficie del yacimiento hasta que quedase como la encontramos. La administración de la reserva natural nos había ayudado en este aspecto durante estos años llevándose en carretillas la cal sobrante que sacábamos de las galerías de la mina. Después de 9 años de trabajo en subsuelo durante los fines de semana





cularmente rico en nódulos de sílex. Los excavadores llegaron a los pocos metros a la primera galería neolítica. Nueve años después, habían localizado 66 pozos mineros, cada uno de los cuales poseía galerías irradiantes cuya zona superficial media era de 25 metros. Se calcula que la región minera de sílex de Rijckholt (Holanda) produjo 41.250 metros cúbicos de sílex, materia prima suficiente como para fabricar unos 153,3 millones de hachas.

podíamos dedicarnos, por fin, a elaborar los resultados.

¿Cómo habían sido las minas de Rijckholt en el Neolítico? Por el descubrimiento de otras dos muestras de ceniza fechables sabemos que el sílex se explotaba allí desde hace algo más de 5000 años. (Las otras dos fechas, 3120 y 3050 a.C., concuerdan con nuestro hallazgo radiocarbónico inicial.) No sabemos durante cuánto tiempo continuó la explotación minera ni cuántos obreros trabajaban. Sin embargo, Joseph Felder ha conseguido sacar algunas estadísticas globales de la operación. En primer lugar, la zona minera de Rijckholt cubre unas 25 hectáreas; de ellas, en sólo 3000 metros cuadrados, encontramos 66 pozos. Si nuestra muestra era representativa, toda la zona de Rijckholt debió tener unas 5000 minas. (Cada pozo con sus galerías adyacentes se considera como una sola mina.) Según los cálculos de Felder, las 5000 minas de Rijckholt produjeron 41.250 metros cúbicos de nódulos de sílex en total, lo que supondría materia prima suficiente para 153,3 millones de hachas. Si calculamos que las minas de Rijckholt estuvieron en actividad a lo largo de unos 500 años, el porcentaje medio de producción habría sido de 1500 hachas diarias, suponiendo que tanto los mineros como los artesanos trabajasen todo el año y no sólo en algunas estaciones.

La profundidad de los pozos que bajaban desde la meseta hasta el estrato calizo con vetas de sílex, variaba desde 10 hasta 16 metros, según el grosor de la arena, grava y loess depositados sobre la caliza; el diámetro oscilaba entre metro y metro y medio. Mientras excavaban en esta zona de relleno suelto, los mineros se protegían de los derrumbamientos construyendo muros de contención con ramas de árbol trenzadas; en la pared del único pozo que limpiamos desde el fondo hasta la superficie pudimos observar huecos con huellas de dichas ramas. (Nosotros mismos nos protegíamos contra derrumbamientos colocando chapas de acero a medida que avanzábamos.) Las galerías que los mineros de Rijckholt habían excavado en el fondo de cada pozo, tenían una superficie total de 25 metros cuadrados, una altura de 60 centímetros, y su anchura variaba según la situación de las vetas de sílex. El método de extracción empleado era el que hoy llamamos de "cámara y pilar": se sacaban bloques enteros de unos 30 centímetros de grosor de la capa caliza que contenía el sílex; otras veces se dejaba intacta para que sirviese como pilar de sujeción del techo de la galería. Para reforzar esos pilares naturales, los mineros rellenaban también la zona minada con caliza sobrante. Una vez agotada una galería, servía para acumular en ella los residuos de la inmediata, y descubrimos que el 95 por ciento de ellas se habían rellenado de este modo.

¿Cómo se las arreglaban los mineros para abrirse paso en esta inmensa masa caliza? Trabajaban con las hachas que los artesanos hacían mientras tanto en el Gran Taller de la superficie. A juzgar por algunas huellas o vacíos que encontramos en los montones de residuos calizos, las hachas de sílex tenían unos mangos de madera de unos 80 centímetros de largo; el tamaño de la hoja en sí era bastante uniforme, oscilando entre 15 y 18 centímetros de largo, 5 de ancho y 3,5 de grosor.

El excavar la caliza era un trabajo duro que requería gran habilidad. Calculamos que un minero gastaría 5 hojas de hacha para excavar un metro cúbico de caliza. Cuando se gastaban, los mismos mineros las afilaban con piedras calizas que hacían de martillo para retocar los bordes; para esto usaban un tipo especial de caliza dura recristalizada. La mayoría de estos "martillos" los encontramos junto a grupos de 10 o 20 hachas en lo que denominamos depósitos de hachas, situados por lo general cerca de un pozo y con muchas lascas de residuos de sílex.

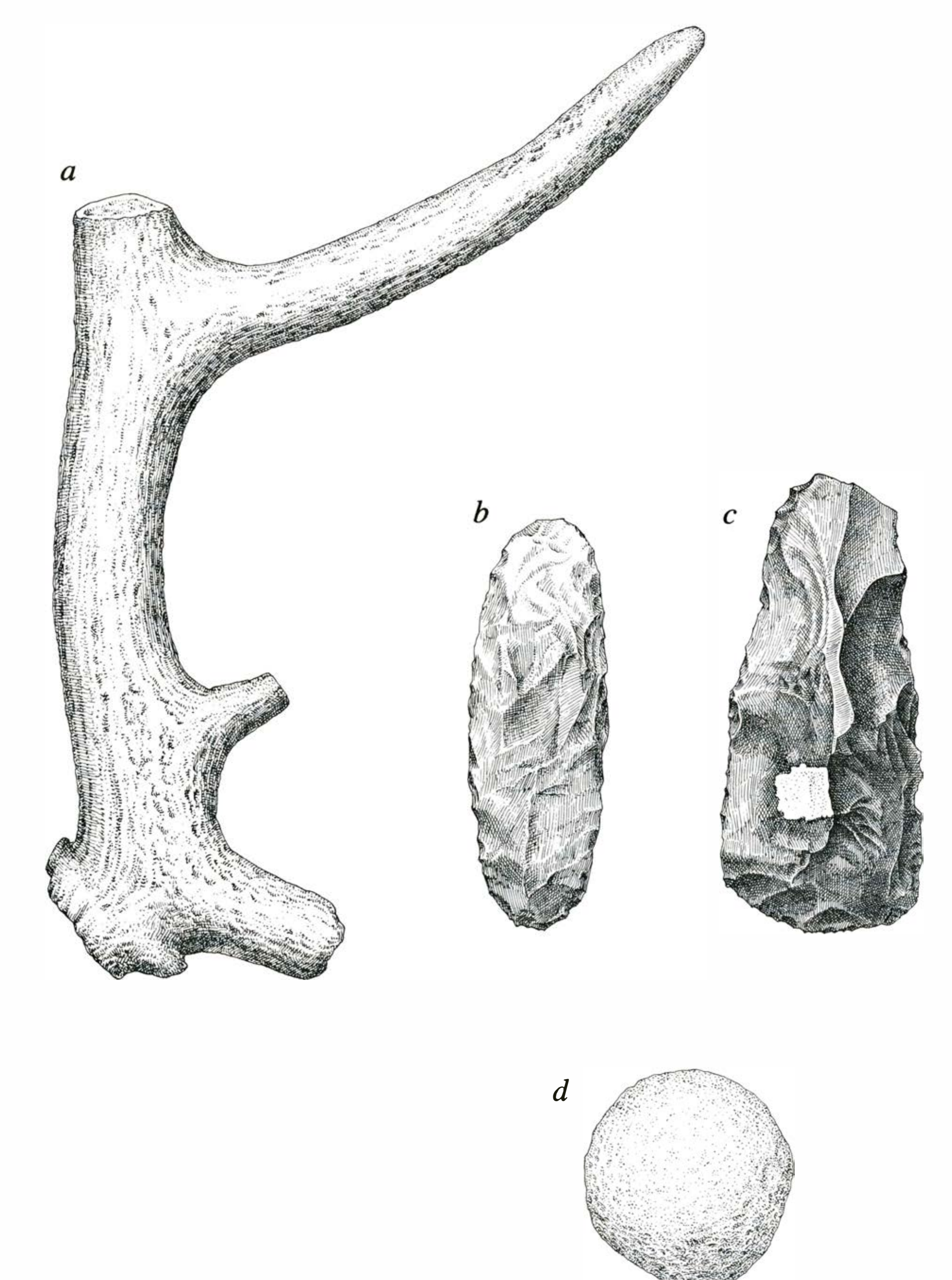
Hallamos en las minas más de 15.000 instrumentos, la mayoría de ellos hojas de hachas rotas o gastadas. Esto quiere decir que, de acuerdo con las previsiones estadísticas de Joseph Felder, quedan en las galerías inexploradas de Rijckholt dos millones y medio de hojas de hacha abandonadas por los mineros. Si es así, estos reinvirtieron menos de un dos por ciento de su producción total de sílex en el trabajo de extracción, con un coste por mina de unas 350 hachas. También encontramos algunos picos de asta parecidos a los que encontró de Puydt en Rijckholt en superficie durante sus investigaciones en el siglo pasado. El hecho de que no encontrásemos muchos indica que no se solían usar para excavar galerías porque la caliza de allí era más dura que el asta.

Parece ser que los mineros de Rijckholt explotaban a la vez más de una mina. Esta suposición se basa en nuestro descubrimiento de "galerías de salida", una especie de pasillos rellenos

sólo en parte con caliza, o vacíos. Las galerías estaban conectadas con dos o más pozos adyacentes. La apertura de tales conexiones no les exigía a los mineros un trabajo excepcional, pues la distancia entre pozos era muy pequeña, probablemente para facilitar el transporte de los nódulos de sílex desde las galerías. Al parecer, la última galería de cada mina conectaba con el pozo de la mina siguiente. Estas galerías de salida se usaban con frecuencia, ya que los residuos de caliza que había en el suelo estaban muy pisoteados y en las paredes y techo encontramos huellas del paso frecuente de los mineros arrastrándose.

No se sabe con certeza cómo bajaban y subían los mineros desde el fondo del pozo, ni cómo sacaban a la superficie los nódulos de sílex. Parece lógico pensar que usaban cuerdas para ambas funciones, pero la única prueba que disponemos para ello son los surcos profundos que hay en la caliza allí donde terminan los pozos y empiezan las galerías.

No encontramos restos de lamparillas ni de antorchas bajo tierra. Quizá porque su combustión habría consumido el oxígeno que los mineros necesitaban más para respirar que para servir de iluminación. Los depósitos de hachas estaban cerca del fondo de los pozos; ello pudiera justificarse porque los mineros precisaban la luz para afilar las hojas. Sin embargo, parece ser que los que trabajaban en las galerías se las arreglaban con cualquiera que fuese la iluminación indirecta que se filtrase desde la superficie.



INSTRUMENTOS DE MINERO encontrados en las galerías de Rijckholt. Pueden apreciarse: picos de asta (*a*), hachas (*b* y *c*) y martillos para reparar las hachas usadas (*d*). Algunos martillos eran de cuarzo y otros de una caliza muy dura, como el que se recoge arriba. Se gastaban cinco hachas por metro cúbico de caliza removida, lo que elevaba a 350 el número de hachas inutilizadas en cada mina.

En el primer año en que se trabajó en todos los túneles, dimos con la prueba críptica de que los mineros realizaban prácticas de culto. El 5 de noviembre de 1965 encontramos, al final de una galería recién abierta, un cráneo humano sin mandíbula inferior dentro de un hoyo pequeño. Salvo este lugar, el resto de la galería estaba completamente rellena de derrubio calizo. Sin duda habían colocado allí el cráneo, y sólo el cráneo, posiblemente como símbolo de un segundo enterramiento. Examinado por G. van Vark, del Laboratorio de Anatomía y Embriología de la Universidad de Groningen, resultó ser de un hombre de unos 40 años. Presentaba lesiones en la sien izquierda y sobre el ojo izquierdo, pero el hecho de que el hueso se hubiese regenerado indica que las heridas no le produjeron la muerte. No es fácil imaginar a qué práctica responde la presencia de ese cráneo en la mina. Tan sólo podemos suponer que era alguna forma de culto a los antepasados.

En 1975 tuvo lugar en Maastricht el segundo de una serie de simposiums sobre sílex, cuyo programa incluía una visita a Rijckholt. Por entonces, hacía tres años que se había abandonado la excavación, los soportes de acero de los túneles se estaban oxidando y las galerías presentaban signos de deterioro. Los participantes en el simposium presentaron un informe al Ministerio de Cultura holandés señalando que las minas de Rijckholt eran un monumento prehistórico de la mayor importancia, y pedían que el gobierno tomase medidas para su conservación. Tanto el gobierno como varias empresas privadas ofrecieron su ayuda, y a principios de ese mismo año se dictaron normas y se dieron órdenes para preservar la zona excavada.

Se ha profundizado el suelo del túnel para que los visitantes puedan recorrerlo de pie. Se han impermeabilizado el techo y las paredes. A la entrada del túnel, hay un pequeño cobertizo para recibir a los visitantes y se ha iluminado por dentro con luz eléctrica. Los turistas e investigadores pueden ahora venir a ver cómo sus antepasados explotaban el sílex que necesitaban para fabricar instrumentos hace 5000 años. Si se echa la vista atrás, parece sorprendente que 20 arqueólogos aficionados hayan hecho esto posible trabajando los viernes por la noche.

Neutrones ultrafríos

Cuando los neutrones se mueven tan despacio que no logran atravesar una superficie sólida, pueden confinarse en una botella metálica. Su almacenamiento permitiría medir algunas propiedades fundamentales

R. Golub, W. Mampe, J. M. Pendlebury y P. Ageron

Está arraigada la idea de que el progreso en el conocimiento de las partículas elementales de la materia procede sólo del campo de las cada vez más altas energías. Pero también a bajas energías hay campos ignotos por explorar. Y así, John G. King, del Instituto de Tecnología de Massachusetts, ha podido decir que queda tanto por aprender entre la energía más baja conseguida y la energía cero como entre la energía más alta hasta ahora alcanzada y la energía infinita.

La avanzadilla en la frontera de la baja energía está formada por los neutrones con una energía de 10^{-7} electronvolt. Un electronvolt es la energía adquirida por un electrón que se acelera por una diferencia de potencial de un volt. Los neutrones tienen, pues, la energía que un electrón recibiría en virtud de una diferencia de potencial de una diezmillonésima de volt, es decir, 0,1 microvolt. Esta es la energía media de una partícula en un gas cuya temperatura fuera de una milésima de grado Kelvin, es decir, una milésima de grado Celsius sobre el cero absoluto. En este intervalo de energía, los neutrones se parecen a un gas de este tipo, y se les denomina neutrones ultrafríos. No es fácil hacerse una idea de cuán pequeña es su energía. Mientras que las partículas de alta energía tienen velocidades cercanas a la de la luz, los neutrones ultrafríos se mueven sólo a unos pocos metros por segundo (la velocidad a que se desplaza un atleta). Más aún, sólo pueden elevarse alrededor de un metro del campo gravitacional de la Tierra antes de pararse y volver a caer.

Una propiedad sorprendente de los neutrones ultrafríos es que sufren una reflexión total por las superficies sólidas. A energías más altas, los neutrones penetran fácilmente la materia, pero cuando se mueven con suficiente lentitud, rebotan en la mayoría de las superficies materiales. Por cuya razón, pueden almacenarse en una botella hecha de materiales ordinarios, como un metal o un vidrio. También pueden confinarse mediante un campo magnético. La mayoría de las otras partículas elementales deben ser estudiadas en vuelo; en cambio, podemos examinar los neutrones ultrafríos casi en reposo.

La posibilidad de confinar neutrones ultrafríos durante largos períodos sugiere dos de sus aplicaciones: en primer lugar, medir la vida media del neutrón, que se conoce con una precisión de sólo un 1,5 por ciento, aproximadamente. En segundo lugar, hallar el momento dipolar eléctrico del neutrón, que es una magnitud de gran interés teórico. Los neutrones ultrafríos pueden servir también como instrumento para la investigación de otros sistemas físicos, tales como películas delgadas o superficies sólidas.

Los neutrones ligados en los núcleos atómicos constituyen más de la mitad de la masa terrestre, pero los neutrones libres son raros en la naturaleza. La razón estriba en que, fuera de los núcleos, los neutrones son partículas inestables: a través de un proceso denominado desintegración beta, un neutrón se desintegra en un protón, un electrón y una partícula sin masa llamada neutrino. La vida media de un neutrón es levemente superior a los 15 minutos.

Los neutrones libres empleados para hacer experimentos se obtienen en gran abundancia en un reactor nuclear, donde se desprenden como productos secundarios de la fisión nuclear. Los neutrones son emitidos con energías de varios millones de electronvolt, pero pronto pierden la mayor parte de ésta como resultado de sus colisiones con los núcleos atómicos del material que rodea el núcleo del reactor, que se llama moderador. Tras pocas colisiones, la energía media de los neutrones es comparable a la energía asociada con los movimientos térmicos de los núcleos del moderador. Los neutrones están, entonces, en equilibrio térmico con los núcleos, es decir, pueden considerarse como un gas a la misma temperatura que el moderador. Si el moderador se halla a la temperatura ambiente (alrededor de 300 grados Kelvin), el gas de neutrones tiene la misma temperatura y la energía media de los neutrones es de unos 25 milielectronvolt.

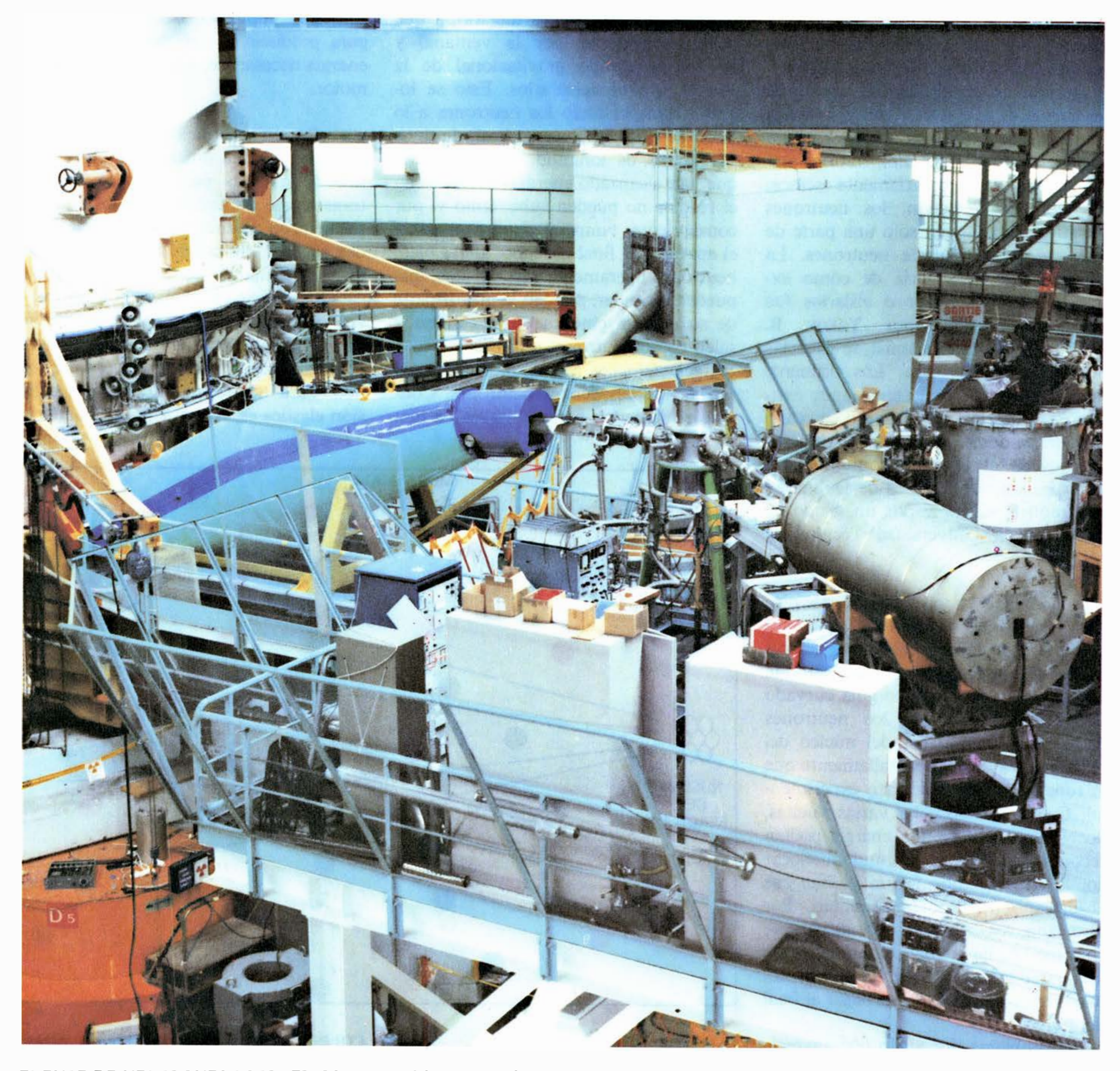
La temperatura del gas de neutrones define no sólo la energía media, sino también el intervalo sobre el que se extienden las energías. A la temperatura ambiente hay neutrones con energías sustancialmente superiores e inferiores a los 25 milielectronvolt; aunque su número disminuye a medida que se alejan de la energía media. A la temperatura del hidrógeno líquido (20 grados Kelvin), la energía media es menor y también lo es el intervalo de energías.

Como las energías se hallan distribuidas en un intervalo comparable a la energía media, podemos encontrar neutrones con energía tan pequeña como se desee a la temperatura ambiente, si bien constituyen una fracción muy pequeña del flujo total. Por definición, los neutrones situados cerca del máximo de la distribución energética a la temperatura ambiente, que tienen una energía de unos 25 milielectronvolt, son llamados neutrones térmicos. Conforme su energía disminuye, se denominan neutrones fríos, muy fríos y ultrafríos. Las fronteras entre una y otra categoría suelen ser un tanto vagas, pero hay un criterio claro para definir los neutrones ultrafríos: la propiedad de la reflexión total.

La reflexión total de los neutrones por las superficies sólidas es análoga a la reflexión interna total de la luz dentro de un prisma de vidrio. Y la mejor forma de entenderla es considerar los neutrones como ondas de materia. En mecánica cuántica, toda partícula puede representarse como una onda cuya longitud es inversamente proporcional a la raíz cuadrada de la energía de la partícula. Para un neutrón térmico con una energía de 25 milielectronvolt la longitud de onda es 1,8 angstrom, un angstrom es

igual a 10^{-8} centímetros. La distancia es comparable al espaciado entre los átomos de un sólido. Para un neutrón ultrafrío con una energía de 10^{-7} electronvolt la longitud de onda es de 900 angstrom. Para la longitud de onda más corta un sólido es una red de núcleos atómicos con espacios vacíos entre ellos, por los que penetra fácilmente la onda. Para longitudes de onda mayores, el sólido actúa como una barrera continua, y la amplitud de la onda disminuye rápidamente después de atravesar la superficie, anulándose en unos 100 angstrom. La probabilidad de encontrar el neutrón en un cierto punto está relacionada con el cuadrado de la amplitud de la onda en este punto; por tanto, es improbable encontrar el neutrón en el sólido. Es repelido por la superficie como un todo. (Este efecto fue predicho y demostrado por Enrico Fermi, en 1945.) El neutrón puede vencer la repulsión y penetrar en el material sólo si su velocidad tiene una componente normal a la superficie que sobrepase cierto valor crítico.

La velocidad crítica que un neutrón necesita para poder penetrar una sustancia es una propiedad de ésta. Para niquel, hierro, berilio y cobre es de alrededor de unos seis metros por segundo, para la mayoría de los otros materiales es sustancialmente menor. En una superficie de cobre, un neutrón que se mueva a 600 metros por segundo será siempre reflejado sólo si incide sobre la superficie con un ángulo menor que aproximadamente medio grado, para que la velocidad normal a la superficie sea menor que el valor crítico. Las reflexiones rasantes de este tipo se usan en espejos y en tubos



FUENTE DE NEUTRONES ULTRAFRIOS: se trata del reactor nuclear del Instituto Max von Laue-Paul Langevin en Grenoble, Francia. El reactor se encuentra situado detrás de la pared curva de la izquierda. Los neutrones salen por un tubo guía inclinado, que está rodeado por un gran apantallamiento. Desde el tubo guía, los neutrones pasan por una válvula de distribución (montada sobre el trípode verde) al aparato experimental. Cuando se hizo la fotografía, se estaban preparando dos experimentos. El cilindro con el eje vertical contiene un anillo de almacenamiento magnético, construido por Wolfgang Paul y Uwe Trinks, de la Universidad de Bonn. Se usará para medir la vida media del neutrón. El cilindro horizontal, más grande, es parte de un experimento que están realizando Norman F. Ramsey, de la Universidad de Harvard, Kenneth F. Smith, de Sussex, los autores (Golub, Mampe, Pendlebury y Ageron) y otros físicos. El experimento se centra en la búsqueda de un momento dipolar eléctrico para el neutrón.

de guía que pueden transportar haces de neutrones con sólo una pequeña pérdida de intensidad.

Los neutrones ultrafríos son aquellos cuya velocidad total es menor que la velocidad crítica para un material dado. Es seguro, por tanto, que serán reflejados por la superficie para cualquier ángulo de incidencia. Para confinarlos, bastará con colocarlos en una botella cerrada cuyas paredes sean del material apropiado. Precisando más, la velocidad crítica para la reflexión total se puede establecer en los 6 metros por segundo, velocidad crítica característica del cobre. Un neutrón que se mueve a 6 metros por segundo tiene una energía de 2×10^{-7} electronvolt, que corresponde a una temperatura de dos miligrados Kelvin.

Aunque un cálculo directo, a partir de las curvas de equilibrio térmico, prevé la existencia de neutrones ultrafríos en un reactor nuclear a la temperatura ambiente, son extraordinariamente escasos. A 300 grados Kelvin, los neutrones ultrafríos constituyen sólo una parte de 10^{11} del flujo total de neutrones. La primera propuesta seria de cómo extraerlos de un reactor o aislarlos fue hecha, en 1959, por Yakov B. Zel'dovich, de la Academia de Ciencias de la Unión Soviética. Los primeros experimentos con éxito empleando neutrones ultrafríos se llevaron a cabo simultáneamente, en 1968, por investigadores en el Instituto conjunto para Investigaciones Nucleares de Dubna en la Unión Soviética y por un grupo de físicos de la Universidad Técnica de Munich.

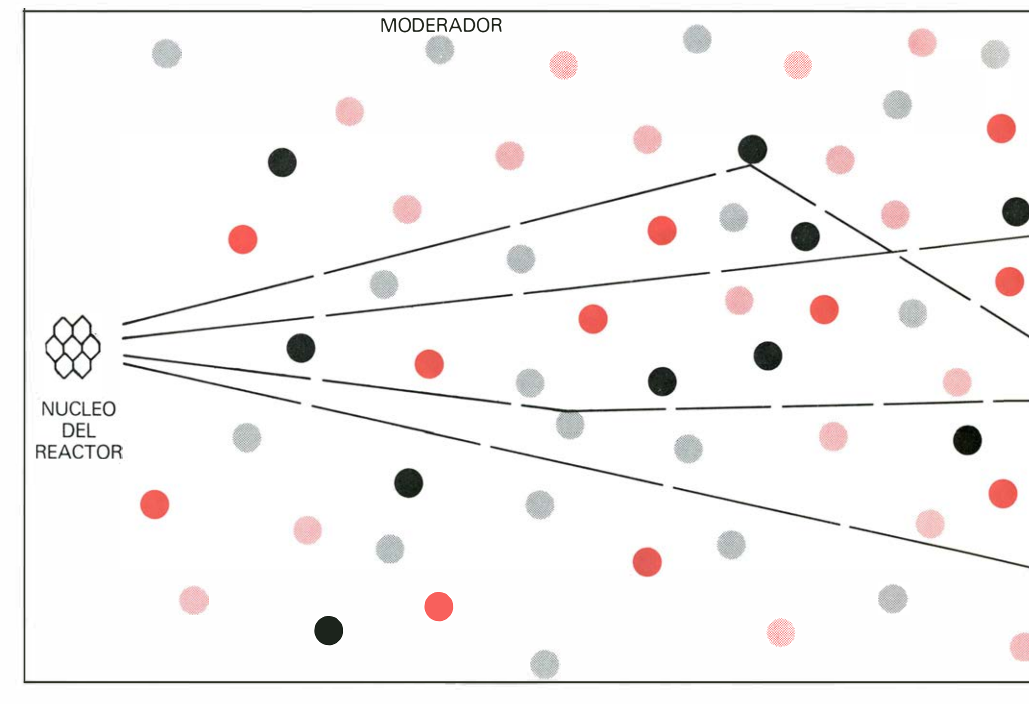
En Dubna, el ya fallecido F. L. Shapiro y sus colegas extrajeron neutrones de un reactor de baja potencia y demostraron que podían almacenarse durante varios segundos. Un tubo guía curvado y horizontal conducía los neutrones desde las proximidades del núcleo del reactor, a través del apantallamiento que lo rodeaba, hasta el aparato experimental. Como el tubo daba varias vueltas, pocos neutrones de alta energía podían recorrerlo en toda su longitud; al chocar con la pared del tubo, la atravesaban y se absorbían en el apantallamiento. Por otro lado, los neutrones ultrafríos, que eran totalmente reflejados, rebotaban de una pared a otra y tenían una probabilidad razonablemente alta de llegar al final del tubo.

Un problema inmediato que aparecía en el diseño era el siguiente: había que hacer el vacío en el tubo guía; por tanto, debía cerrarse en el extremo del interior del reactor. La ventana metálica podía ser muy delgada, pero aun así la reflexión total excluía del tubo precisamente los neutrones deseados, los ultrafríos. El grupo de Shapiro venció esta dificultad instalando dentro de la ventana un material, un convertidor. Algunos de los neutrones que se movían con velocidad suficiente para atravesar la ventana cedían parte de su energía a los núcleos del convertidor y se transformaban, consiguientemente, en neutrones ultrafríos. Los mejores convertidores suelen ser materiales ricos en hidrógeno, tales como el polietileno o el agua.

En Munich, Albert Steyerl obtuvo neutrones ultrafríos por un método distinto: seleccionó neutrones que se movieran a velocidades algo mayores que la velocidad crítica (de forma que pudieran penetrar por la ventana) y empleó el campo gravitacional de la Tierra para desacelerarlos. Esto se lograba transportando los neutrones a lo largo de un tubo guía vertical, de unos 11 metros de longitud. Los neutrones que han alcanzado energías ultrafrías en el reactor no pueden subir tanto y, por consiguiente, tampoco pueden alcanzar el aparato del final del tubo; ahora bien, neutrones ligeramente más rápidos, que pueden ascender, pierden la mayor parte de su energía cinética en el camino y salen como neutrones ultrafríos.

Recientemente, Steyerl ha ideado otro método de desacelerar neutrones. Consiste en hacerlos pasar a través de una "turbina" que absorbe parte de su energía. La turbina es una rueda de 1,7 metros de diámetro con unas 660 superficies reflectantes curvadas y fijas a su circunferencia. Se hace girar la rueda con una velocidad tangencial de 25 metros por segundo, y se mandan neutrones que van a unos 50 metros por segundo de forma que incidan sobre las paletas con un ángulo rasante. Cada neutrón puede ser reflejado varias veces desde la superficie cóncava de la paleta, y en cada rebote cede energía que produce el movimiento de la turbina, exactamente igual que el agua que cae o los gases en expansión. Los neutrones son, por supuesto, demasiado débiles para producir el giro de la turbina; la energía necesaria es suministrada por un motor.

La reflexión total de los neutrones ultrafríos es total únicamente en el sentido de que las partículas no pueden transmitirse a través del sólido. Cuando un neutrón alcance una superficie pueden suceder tres cosas. En primer lugar, que se dé una colisión elástica; en cuyo caso el neutrón es reflejado por la superficie sin sufrir cambio en su energía o velocidad, en otras palabras, sin intercambiar energía con los núcleos del sólido. Para neutrones ultrafríos, la colisión elástica es, con mucho, el resultado



COLISIONES DE NEUTRONES con núcleos atómicos. Tales choques dan origen a neutrones con un ancho espectro de energías; entre ellos: unos cuantos neutrones ultrafríos. Los neutrones son emitidos en el núcleo del reactor como producto secundario de la fisión de los núcleos de uranio. Salen con energías de varios millones de electronvolt y velocidades del orden del 10 por ciento de la velocidad de la luz. La mayor parte de

más probable de un choque, y si fuera el único resultado posible, los neutrones podrían ser confinados indefinidamente (al menos hasta que ocurriera su desintegración).

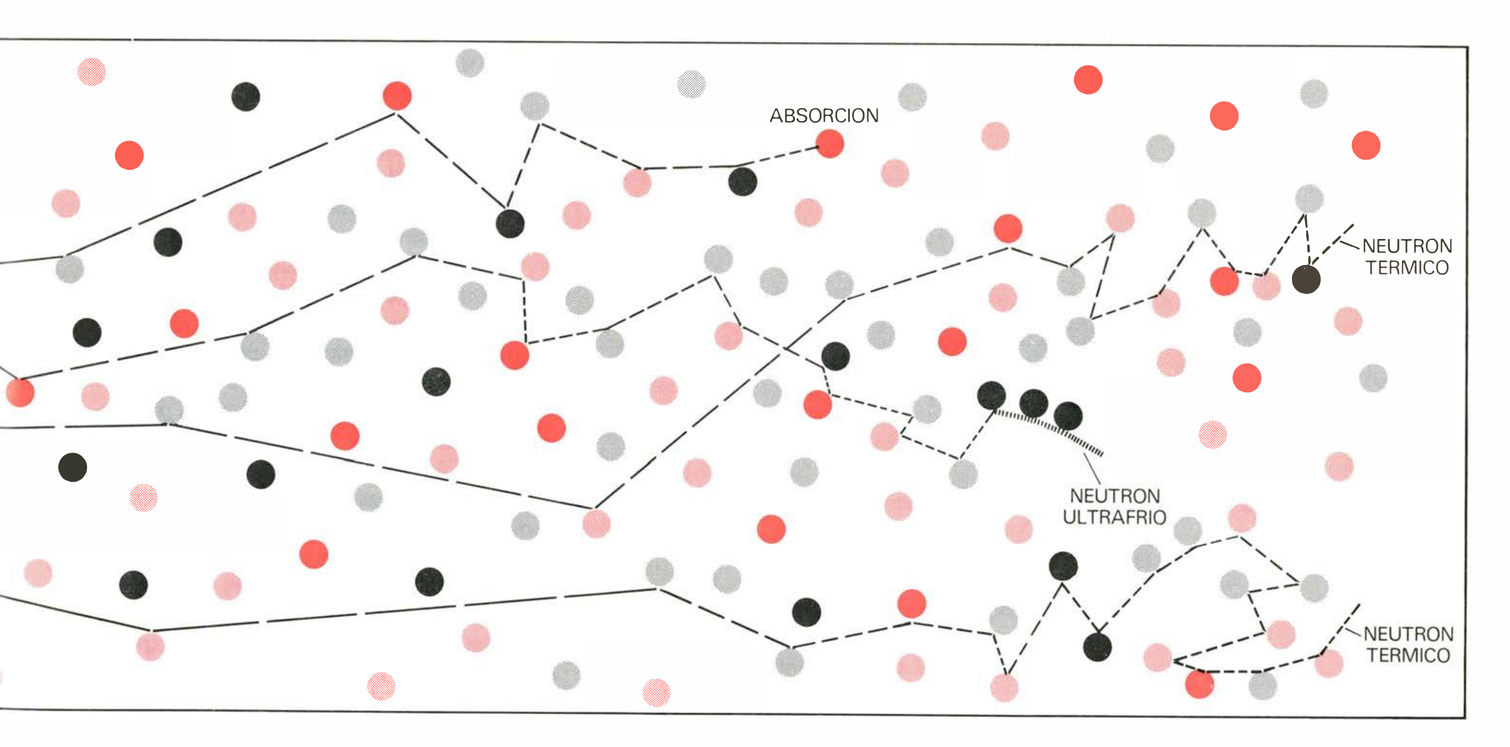
Otra posibilidad es que se dé una colisión ineslástica. En ella hay intercambio de energía entre el neutrón y el núcleo. Debido a las excitaciones térmicas, los núcleos de un sólido vibran alrededor de sus posiciones de equilibrio. Como el neutrón penetra a lo largo de una pequeña distancia en el material, se solapa con algunos de los núcleos y puede así interaccionar con ellos. Si el neutrón interacciona con un núcleo que está retrocediendo, puede perder energía; si el núcleo se está aproximando en el instante de la colisión, el neutrón puede ver incrementada su energía. En cualquier caso, el neutrón sigue siendo reflejado por la superficie, pero con una velocidad distinta. Es mucho más probable que ocurra un aumento de energía que una pérdida de la misma, y un neutrón que gane energía excederá la velocidad crítica. Luego, cuando golpee de nuevo la pared de la vasija confinante, se transmitirá en vez de reflejarse, perdiéndose definitivamente de entre la población confinada.

La tercera posibilidad con que cuenta el neutrón es que sea absorbido por un núcleo de la pared. Así, un núcleo de carbono que tiene seis protones y seis neutrones, con una masa atómica de 12, puede absorber un neutrón adicional y convertirse en un núcleo del isótopo carbono 13. Como en el caso de las colisiones inelásticas, la absorción elimina definitivamente el neutrón de la población.

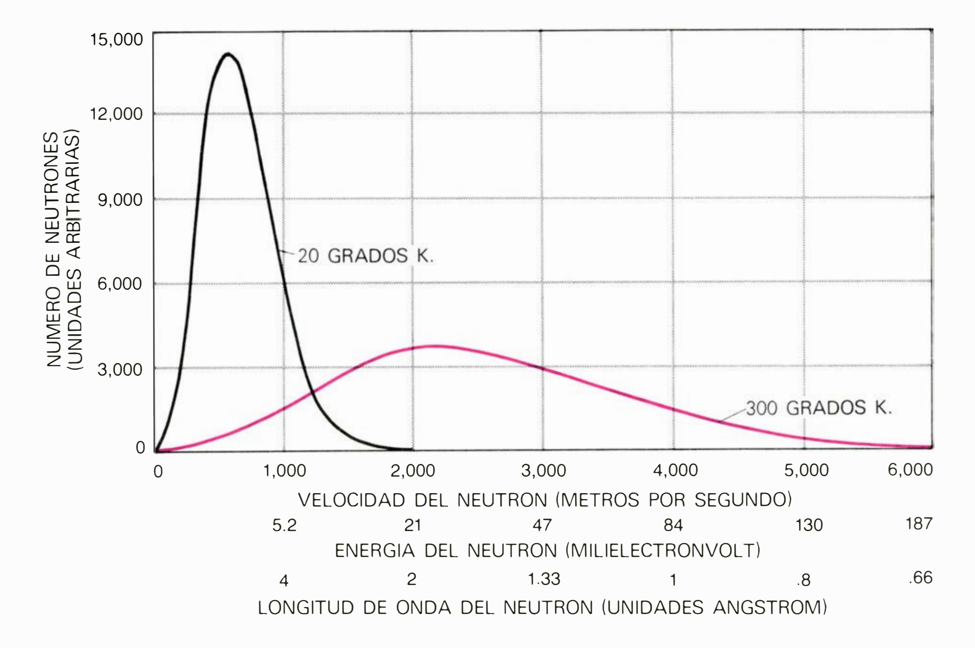
La probabilidad de que un neutrón se pierda por absorción o por colisión inelástica depende del material de las paredes. Las colisiones inelásticas son más probables con núcleos ligeros, porque a una temperatura dada se mueven más deprisa que los pesados. La probabilidad de absorción depende de forma compleja de la composición del núcleo, pero esta probabilidad ha sido medida para todos los núcleos más usuales. A partir de las propiedades del material de las paredes podemos calcular el número medio de veces que el neutrón llegará a chocar con la pared antes de perderse a través de uno de estos mecanismos. Para el cobre, el número calculado es de alrededor de 6000; para el berilio, de 200.000.

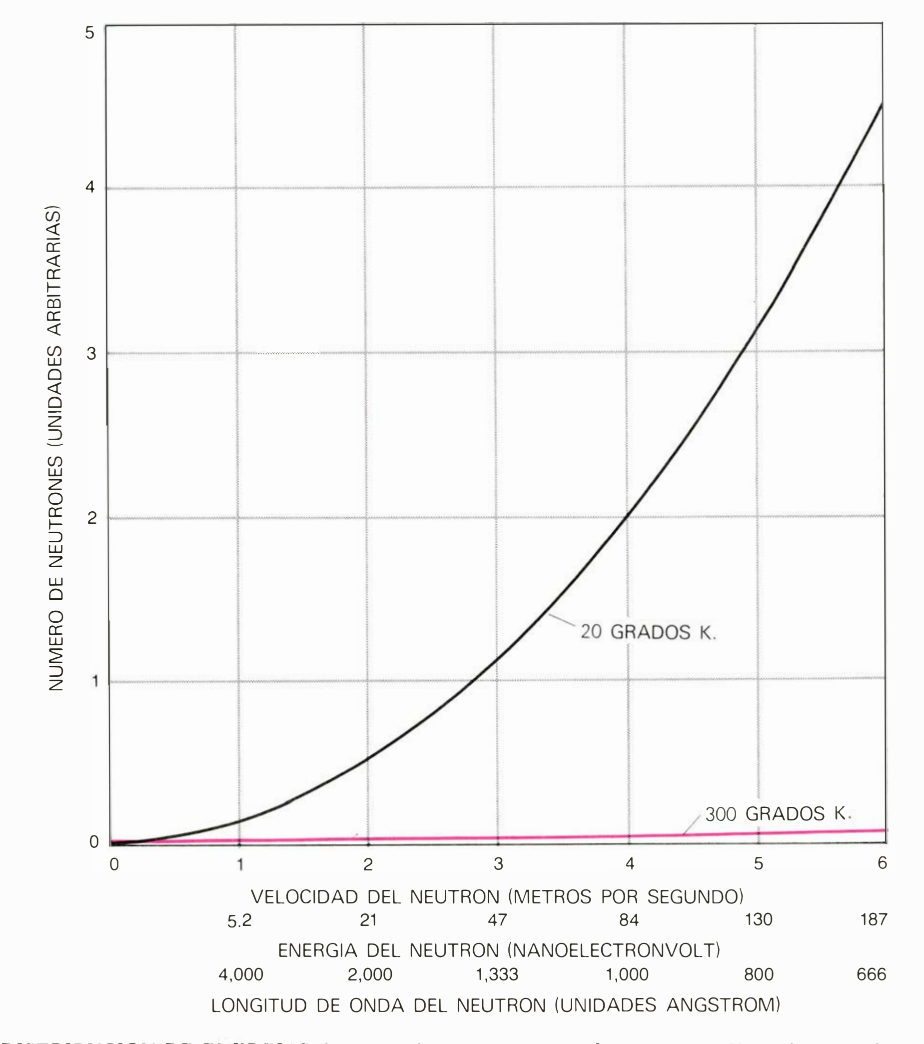
Para neutrones colocados en un recipiente cerrado el tiempo esperado de almacenamiento es directamente proporcional al número medio de veces que un neutrón choca con las paredes antes de perderse. (El tiempo de almacenamiento depende también del tamaño y forma de la vasija y de la velocidad de los neutrones, ya que estos factores determinan el tiempo entre los choques con las paredes.) Un hecho preocupante es que los tiempos de almacenaje medidos son significativamente menores que lo que se había esperado. Una larga serie de experimentos, hechos en varias instituciones con distintas técnicas y aparatos, han confirmado todos esta anomalía. Los resultados de estos experimentos pueden resumirse de una forma sencilla: a pesar de las grandes diferencias entre las probabilidades teóricas de pérdida, en la mayoría de los materiales se encuentra el mismo número medio de choques con las paredes antes de que los neutrones se pierdan, que es aproximadamente 1000. Esta anomalía, descubierta ya hace diez años, no acaba de entenderse bien.

La hipótesis más aceptada atribuye las excesivas pérdidas de neutrones a impurezas sobre la superficie de las paredes, probablemente iones hidroxilo (OH^-), u otros grupos químicos o sustancias ricas en hidrógeno. Un núcleo de hidrógeno puede absorber un neutrón (formando un núcleo de deuterio), pero lo que es más importante, el hidrógeno es el más ligero de todos los núcleos, pues consta de un solo protón, y, por consiguiente, sus movimientos térmicos son excepcionalmente rápidos. Una capa de hidró-



esta energía es cedida rápidamente al material que rodea el núcleo, que recibe el nombre de moderador. En el Instituto Laue-Langevin el núcleo está sumergido en agua pesada (D_2O), que sirve como moderador. Tras unas pocas colisiones, la energía de los neutrones es comparable a la asociada con los movimientos térmicos de los núcleos del moderador; los neutrones pueden entonces considerarse como un gas en equilibrio térmico con el moderador. Los núcleos tienen también energías distribuidas en un cierto intervalo; cuando un neutrón colisione con un núcleo que se mueva de un modo particularmente rápido, es probable que el neutrón gane energía. Por contra, en la colisión con un núcleo lento, el neutrón cederá generalmente energía. Después de muchas colisiones, la energía media del neutrón alcanza un valor estacionario determinado por la temperatura del moderador, si bien algunos neutrones están más calientes y otros más fríos. Los neutrones ultrafríos se producen sólo en aquellos sucesos improbables en los que el neutrón pierde casi toda su energía en una única colisión. Como tales sucesos son improbables, los neutrones ultrafríos constituyen una componente pequeña del flujo total. Las energías térmicas de los núcleos se representan aquí por sus colores, según la clave interpretativa de la izquierda.





DISTRIBUCION DE ENERGIAS de un gas de neutrones en equilibrio térmico. Viene determinada por su temperatura. En la gráfica superior se muestra el espectro completo de energías de los neutrones, con el número de ellos dados en función de tres cantidades relacionadas entre sí: velocidad, energía y longitud de onda. A la temperatura ambiente (300 grados Kelvin) la velocidad media es de unos 2350 metros por segundo, pero la distribución es muy ancha. A la temperatura del hidrógeno líquido (20 grados K) el intervalo es mucho más estrecho y de unos 600 metros por segundo la velocidad media. En la gráfica inferior se da la porción del espectro correspondiente al extremo de baja energía aumentado por un factor 1000. A los neutrones de esta región, con velocidades inferiores a los seis metros por segundo, se les denomina ultrafríos. Incluso a 20 grados K los neutrones ultrafríos son raros, y a 300 grados constituyen menos de una parte en 10^{11} del flujo total de neutrones. A pesar de lo cual puede extraérseles de un reactor que se halle a la temperatura ambiente. (Ilustraciones de Allen Beechel.)

geno adsorbida en la superficie podría aumentar muchísimo la probabilidad de colisiones inelásticas.

Uno de nosotros (Golub), en colaboración con William A. Lanford, de la Universidad de Yale, ha medido el hidrógeno en superficies preparadas de la misma forma que los materiales usados en las vasijas de confinamiento de neutrones ultrafríos. Encontramos hidrógeno suficiente para explicar el ritmo extraordinariamente alto de pérdida de neutrones. En otro trabajo A. V. Strelkov y M. Hetzelt, del Instituto de Física Nuclear de Alma-Alta en la Unión Soviética, rodearon la vasija de neutrones ultrafríos con detectores que medían la energía de los neutrones que se escapaban. Encontraron que los neutrones que abandonaban la vasija tenían energía cerca de la térmica; y no otra cosa había que esperar si los neutrones habían ganado velocidad a través de colisiones inelásticas con las impurezas.

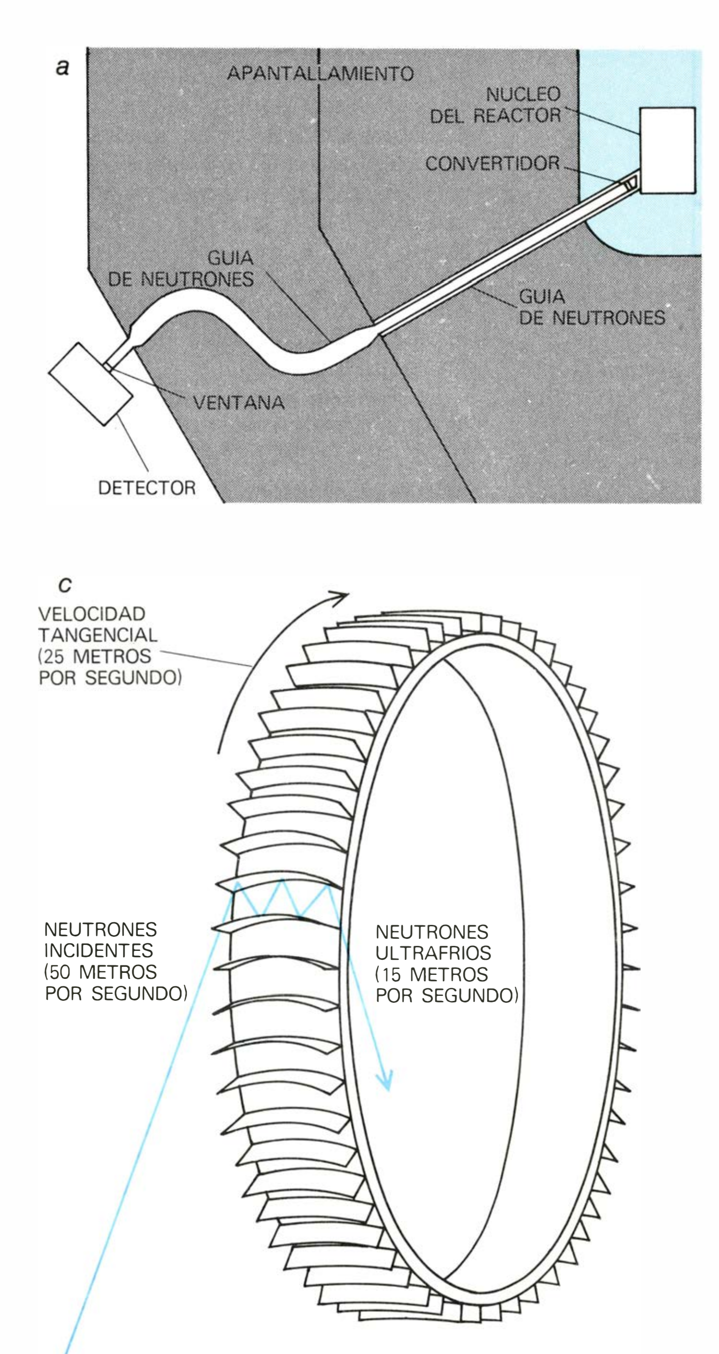
Existen, sin embargo, hechos experimentales que no se reconcilian fácilmente con la hipótesis de las impurezas. Si las impurezas fueran las responsables, el ritmo de pérdida dependería de la temperatura de las paredes. Elevando la temperatura, aumentan los movimientos térmicos de los átomos y algunos de ellos son eliminados totalmente de la superficie, por lo que parece probable algún tipo de cancelación de ambos efectos. Hasta ahora no se ha encontrado ninguna dependencia incontrastable de la temperatura con respecto al tiempo de almacenaje. Los intentos de reemplazar el hidrógeno absorbido por deuterio, que daría una probabilidad de pérdida menor, han fracasado a su vez.

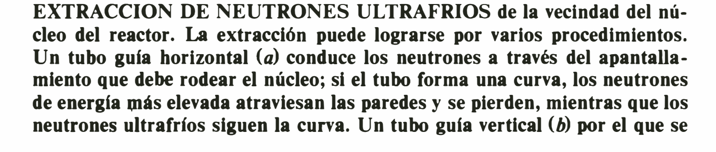
A Strelkov y sus colaboradores debemos un ingenioso experimento en su afán por mejorar el tiempo máximo de almacenaje. Emplearon neutrones con una energía inferior a 3×10^{-8} electronvolt, que es alrededor de 10 veces menor que la energía crítica de muchos materiales. Los neutrones se almacenaban en una vasija cuyas superficies interiores se renovaban continuamente, depositando sobre ellas un vapor metálico. No se detectó ningún aumento en el tiempo de almacenaje cuando se depositaba el nuevo material. Sin embargo, el experimento no es concluyente. El contenido en hidrógeno de las capas recién depositadas no fue medido, y podría no ser apreciablemente menor que el del material de las paredes originales.

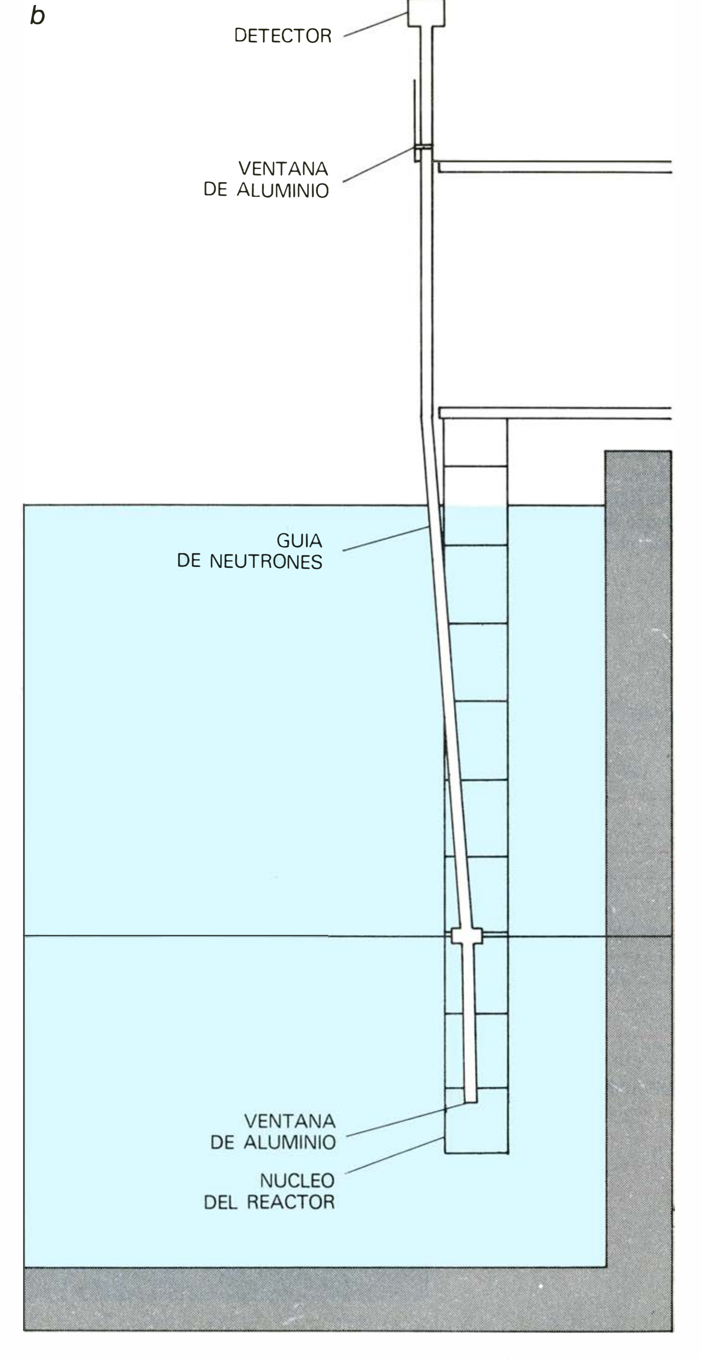
Se han propuesto otros efectos para explicar estos tiempos anormales de almacenaje. Por ejemplo, los neutrones pueden colisionar inelásticamente con átomos que se difunden a través de las

paredes, o pueden escaparse a través de poros en el material de las paredes. Vibraciones de gran escala de las paredes, causadas por ruidos, podrían impartir una velocidad adicional a los núcleos, lo cual aumentaría el ritmo de las colisiones inelásticas. Sin embargo, ninguno de estos efectos parece suficientemente grande para explicar el ritmo de pérdida de los neutrones, y la contaminación de las superficies por hidrógeno continúa siendo la más verosímil hipótesis. Los experimentos que están ahora en fase de preparación, en los que las paredes se enfriarán a la temperatura del helio líquido y la contaminación se reducirá mediante vacíos ultraelevados, prometen dejar claro este punto.

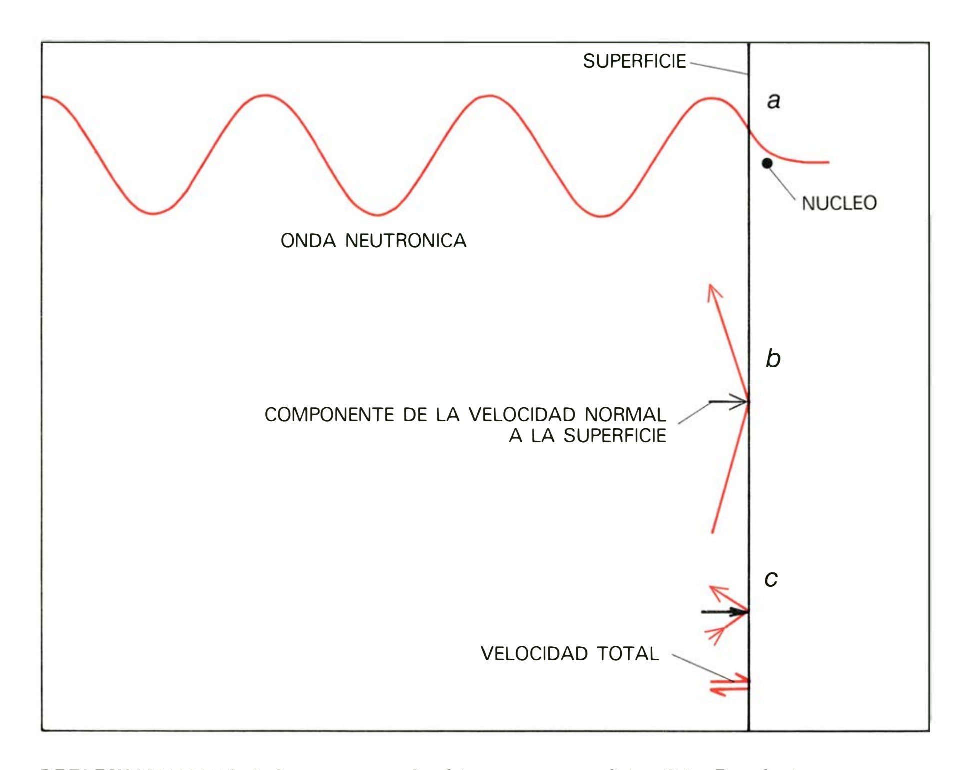
Las complicadas interacciones con la superficie pueden evitarse por completo en otro método de almacenar neutrones ultrafríos, a saber: por confinamiento mediante un campo magnético. El magnetismo es el método normal de controlar las trayectorias de otras partículas elementales, como electrones o protones, en un acelerador de partículas o en un anillo de almacenamiento, pero en estos casos el campo magnético actúa sobre la carga eléctrica de la partícula en movimiento. Al no poseer carga eléctrica, el neutrón no queda afectado por un campo magnético uniforme. Sin embargo, el neutrón tiene un momento dipolar magnético, esto es, tiene un polo magnético norte y sur. El eje del momento dipolar está alineado con el eje alrededor del cual gira la partícula, y



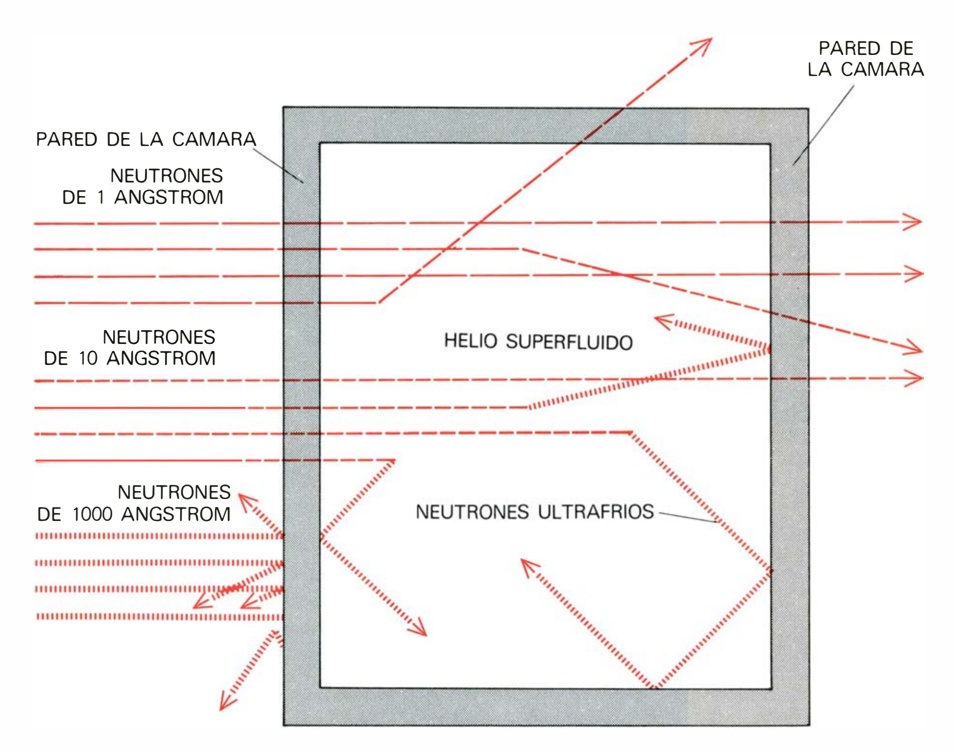




EXTRACCION DE NEUTRONES ULTRAFRIOS de la vecindad del núcleo del reactor. La extracción puede lograrse por varios procedimientos. Un tubo guía horizontal (*a*) conduce los neutrones a través del apantallamiento que debe rodear el núcleo; si el tubo forma una curva, los neutrones de energía más elevada atraviesan las paredes y se pierden, mientras que los neutrones ultrafríos siguen la curva. Un tubo guía vertical (*b*) por el que se introducen neutrones algo más energéticos que los ultrafríos por el extremo inferior, al subir por el aparato, son frenados por el campo gravitacional terrestre y alcanzan la parte superior como neutrones ultrafríos. En una turbina de neutrones (*c*), los neutrones que se mueven a unos 50 metros por segundo colisionan varias veces sobre las paletas que retroceden de una rueda en rápido movimiento, perdiendo energía en cada choque que sufren.



REFLEXION TOTAL de los neutrones ultrafríos en una superficie sólida. Este fenómeno es consecuencia de su interacción con los núcleos del sólido. La onda que describe el movimiento de un neutrón (*a*) tiene una longitud de onda de unos 1000 angstrom (un angstrom es igual a 10^{-8} centímetros), y puede penetrar únicamente unos 100 angstrom en el sólido. Por ello la superficie repele el neutrón, y la repulsión puede ser vencida sólo si el neutrón se mueve hacia la superficie con suficiente rapidez. Con certeza, el neutrón de velocidad moderada (*b*) será reflejado si incide sobre la superficie de forma rasante, de manera que la componente de la velocidad normal a la superficie es menor que la velocidad crítica. Los neutrones ultrafríos (*c*) son aquellos cuya velocidad total no alcanza el valor crítico; por consiguiente salen reflejados, aunque incidan perpendicularmente sobre la superficie. Para el caso del cobre, la velocidad crítica resulta ser de seis metros por segundo; los neutrones que se muevan a velocidad menor que esta última velocidad crítica podrán almacenarse en el interior de una vasija de cobre.



ENFRIAMIENTO DE NEUTRONES en helio líquido. Es un modo de explotar las propiedades mecánico cuánticas de un superfluido para poder confinar neutrones ultrafríos a densidades mayores que las de un reactor. El superfluido posee sólo unos pocos estados posibles de movimiento, que tienen energías muy distintas. Un neutrón con una longitud de onda de unos 10 angstrom puede ceder casi toda su energía al superfluido y es poco probable que gane de nuevo esta energía. Si el superfluido se coloca en una vasija construida de material apropiado, los neutrones de 10 angstrom pueden pasar a través de las paredes, pero al convertirse por acción del helio en neutrones ultrafríos quedan atrapados.

efectivamente el momento angular de spin del neutrón puede ser la razón fundamental de la presencia del momento magnético.

En un campo magnético externo que varíe en magnitud de un punto a otro, el momento magnético dipolar nos suministra un método para condicionar el movimiento del neutrón. En un tal gradiente de campo, un polo debe hallarse sometido a un campo más intenso que el otro polo, y en consecuencia se ejerce una fuerza sobre la partícula. Una configuración de campo apropiada para confinar neutrones es un hexapolo, que se genera mediante seis polos magnéticos, con polaridades norte y sur alternadas, dispuestos alrededor de un cilindro. Las líneas de fuerza que representan la dirección del campo forman una imagen muy complicada, pero la magnitud del campo varía de un modo sencillo. La magnitud es cero a lo largo del eje del cilindro, y en todas las direcciones que se alejan del eje aumenta en la misma razón que el cuadrado de la distancia.

La respuesta del neutrón a un campo hexapolar depende de la orientación del eje de giro y del momento magnético dipolar. Si el momento es paralelo al campo externo, el neutrón se halla sometido a una fuerza que le tiende a apartar del eje del cilindro y es expelido del aparato. Si el momento dipolar es antiparalelo al campo, la fuerza está dirigida hacia el centro. (De acuerdo con las reglas de la mecánica cuántica éstas son las dos únicas orientaciones posibles.) Si un neutrón cuyo dipolo magnético es antiparalelo al campo tiene una velocidad perpendicular al eje demasiado pequeña para vencer la fuerza generada por el gradiente de campo, entonces el neutrón quedará confinado permanentemente. En circunstancias normales, los neutrones están orientados al azar, la mitad paralelos y la otra mitad antiparalelos; por consiguiente, un camo hexapolar puede confinar la mitad de los neutrones de que se dispone.

Con un campo hexapolar cilíndrico, los neutrones pueden escapar por los extremos del cilindro. Se construye una botella magnética idónea para neutrones curvando el cilindro sobre sí mismo para formar un toro. El resultado es un anillo de almacenamiento en donde los neutrones siguen trayectorias helicoidales alrededor del eje del toro. El aparato es totalmente análogo a los empleados en varios laboratorios de física de altas energías para acumular protones o electrones.

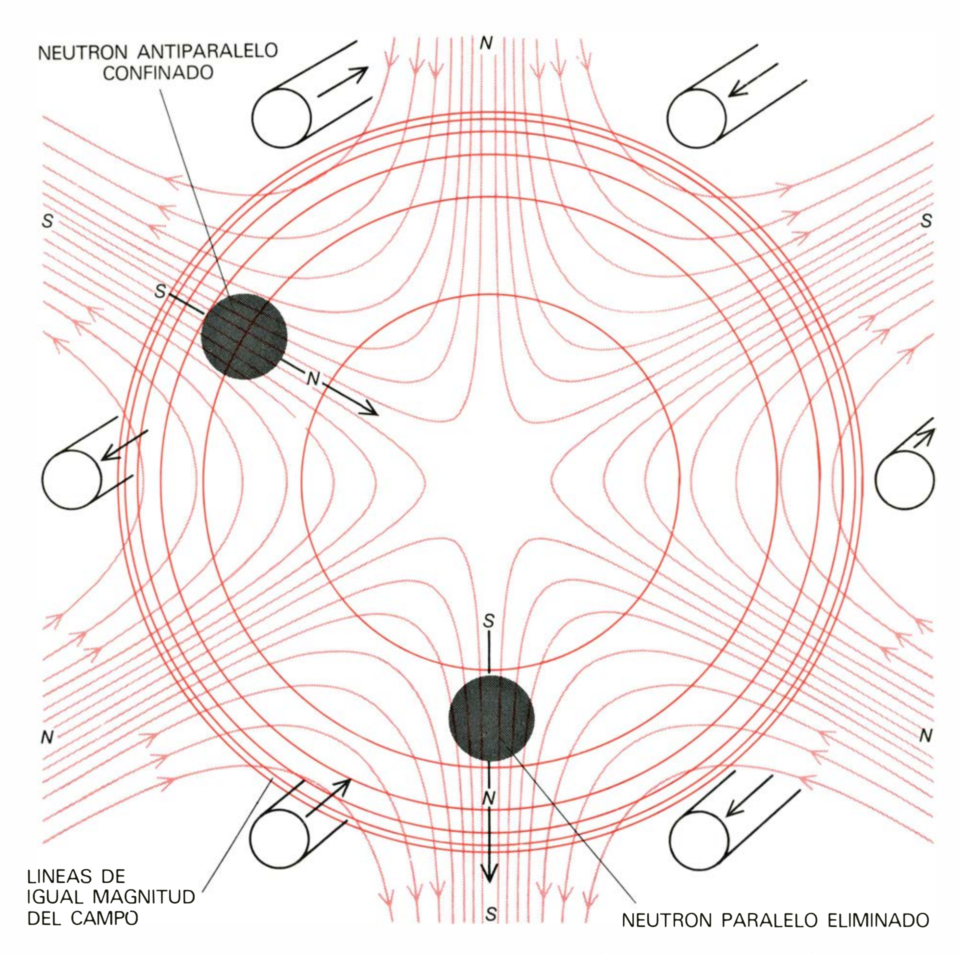
Como ocurría en la vasija material, la botella magnética para neutrones posee una velocidad crítica: se trata de la velo-

cidad radial a la que el neutrón puede vencer el gradiente de campo. La velocidad crítica es proporcional a la raíz cuadrada de la intensidad del campo; en situaciones reales se necesita un campo magnético muy intenso. Wolfgang Paul y Uwe Trinks, de la Universidad de Bonn, construyeron un anillo de esta clase; en él, mediante bobinados superconductores, se obtiene un campo de más de 50.000 gauss. El anillo funciona en el reactor de alto flujo del Instituto Max von Laue-Paul Langerin en Grenoble, un laboratorio compartido por Francia, Alemania Occidental y Gran Bretaña. Se han logrado almacenar neutrones durante períodos de hasta 45 minutos, es decir, alrededor de tres veces la vida media de un neutrón.

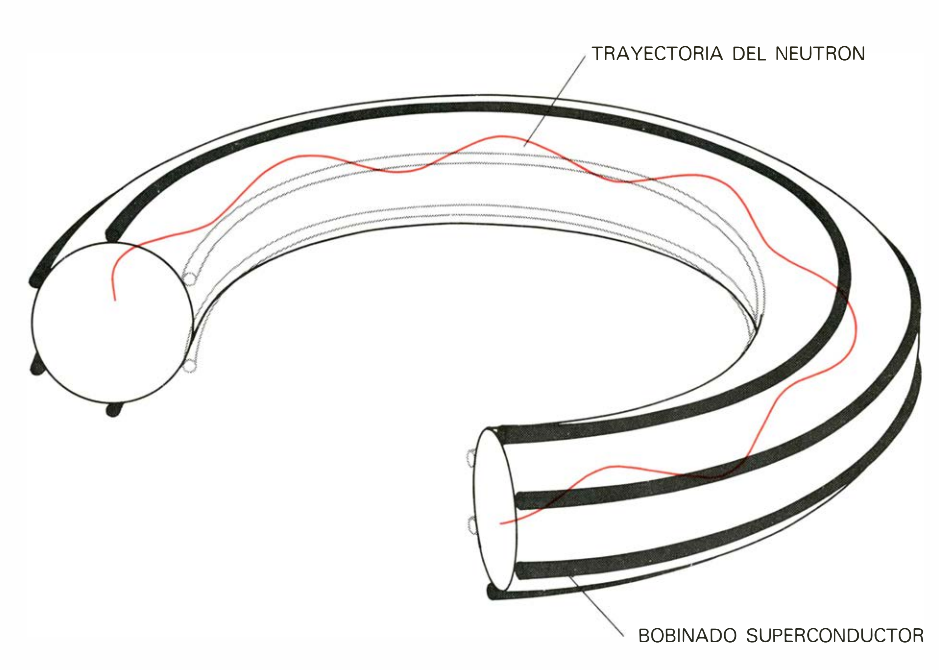
El mismo grupo de investigadores está preparando ahora una botella magnética esférica que se espera que sea aún más eficaz. Tres bobinados superconductores, colocados sobre la superficie de la esfera, uno arrollado según el ecuador y los otros dos paralelos al mismo. El campo creado por estos conductores sale nulo en el centro de la esfera y su magnitud aumenta simétricamente en todas las direcciones que se alejan del centro. Los neutrones de energías suficientemente bajas quedan confinados en órbitas circulares o elípticas análogas a las seguidas por los electrones en un modelo atómico simple.

Incluso con una vasija confinante perfecta habría un límite fundamental para la densidad de neutrones ultrafríos que podrían obtenerse mediante simple extracción de un reactor. La concentración en el aparato no puede exceder a la concentración en el reactor. En realidad, a causa de las pérdidas en los tubos guía y en otros lugares no se puede alcanzar ni siquiera una concentración igual. En este aspecto, el comportamiento de los neutrones vuelve a recordar el de las moléculas de un gas. El sistema formado por el reactor, la botella de neutrones y el tubo guía puede representarse como dos botellas de gas, una llena y otra vacía, conectadas por una tubería. Cuando se abre una válvula en la tubería, la presión en la botella vacía empieza a elevarse gradualmente hasta que iguala a la presión de la botella llena, pero no puede aumentar más.

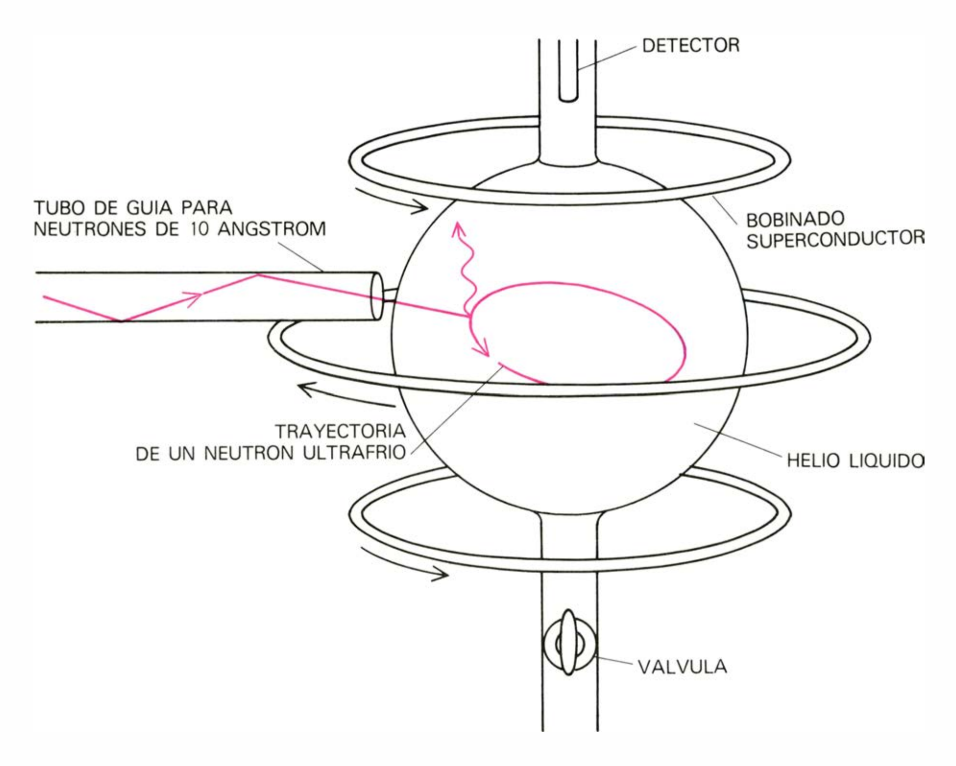
Para alcanzar una densidad mayor de neutrones ultrafríos no hay más alternativa que enfriar los neutrones a su salida del reactor. Los propios métodos de extracción que reducen la energía media de los neutrones, tales como un tubo guía vertical o una turbina, no eliminan calor. Para enfriar un gas no basta redu-



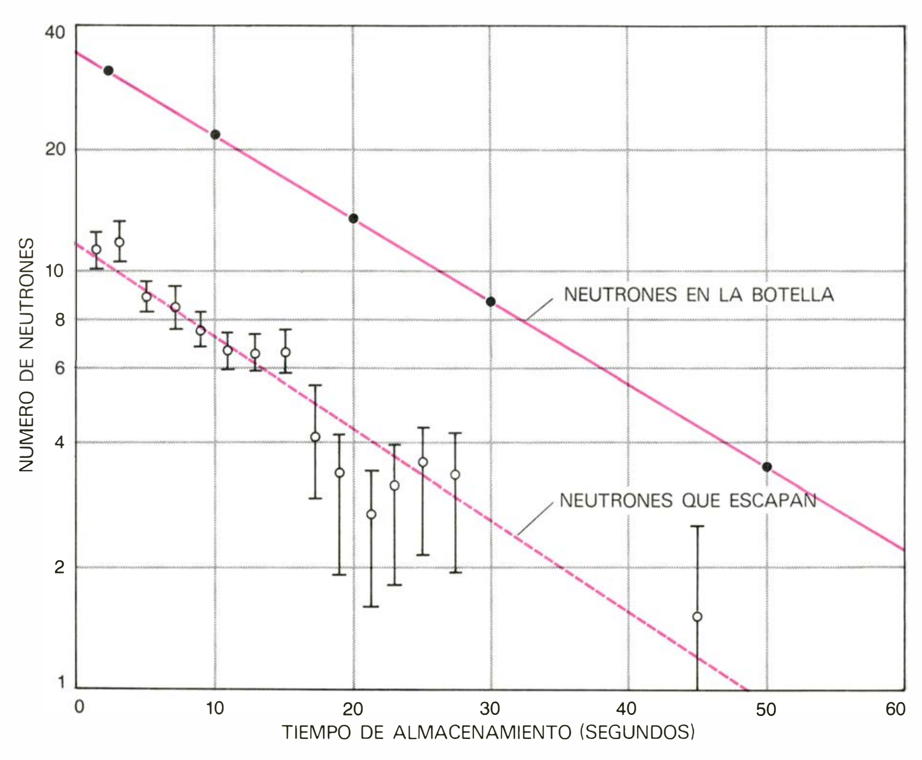
UN CAMPO MAGNETICO HEXAPOLAR también puede confinar los neutrones ultrafríos. Aunque la forma del campo es complicada, la intensidad varía de un modo simple: es nulo a lo largo del eje y aumenta como el cuadrado de la distancia al eje. El gradiente de campo resultante actúa sobre el momento dipolar magnético del neutrón. Un extremo del neutrón está siempre en un campo magnético más fuerte que el otro extremo; por tanto hay una fuerza neta actuando sobre la partícula. La dirección de la fuerza depende de la orientación del neutrón. Un campo cilíndrico hexapolar se genera mediante seis conductores rectilíneos con las corrientes eléctricas circulando en las direcciones de las flechas.



ANILLO DE ALMACENAMIENTO TOROIDAL que atrapa los neutrones ultrafríos mediante un campo magnético hexapolar. Los neutrones describen trayectorias helicoidales a lo largo de la línea central del toro. Los dos bobinados interiores (*en gris*) pueden eliminarse, pues las fuerzas centrífugas evitan que los neutrones escapen en esta dirección. El anillo construido en la Universidad de Bonn y que funciona en Laue-Langevin sigue este esquema. No se muestra el aparato para inyectar los neutrones.



BOTELLA MAGNETICA ESFERICA. En su interior existe un campo magnético que es nulo en el centro y aumenta en todas las direcciones como el cuadrado de la distancia al centro. La vasija, que está en fase de construcción por Paul y sus colaboradores en Bonn, funcionará en el Instituto Laue-Langevin como una fuente superfluida de neutrones ultrafríos. La botella se llenará con helio líquido, se establecerá el campo magnético y a continuación se inyectarán, a través de la pared, neutrones con una longitud de onda de unos 10 angstrom. Las interacciones con helio convertirán estos neutrones en neutrones ultrafríos, que quedarán atrapados por el campo magnético cuando se saque el helio.



TIEMPO DE ALMACENAMIENTO de neutrones ultrafríos en una vasija material; es substancialmente más corto que lo que sugieren los cálculos teóricos. La causa probable de la discrepancia se identificó rodeando una vasija de cobre con detectores de neutrones de forma que se podían contar tanto los neutrones que quedaban en la vasija como los que se escapaban de la misma (es decir, la energía de los neutrones que escapan). El hecho de que ambas cantidades disminuyen al mismo ritmo sugiere que los neutrones ganan energía como resultado de las colisiones inelásticas y se pierden a continuación al atravesar las paredes. El elevado ritmo de colisiones inelásticas está causado probablemente por las impurezas, principalmante hidrógeno, en la superficie de las paredes. El experimento fue llevado a cabo por A. V. Strelkov y M. Hetzelt en el Instituto de Física Nuclear de Alma-Ata, en la Unión Soviética.

cir la energía media de las partículas; el intervalo de energías posibles alrededor de la media debe reducirse también. La guía vertical y la turbina sólo restan una cantidad fija a la energía de cada neutrón, sin cambiar el intervalo.

Una fuente de neutrones ultrafríos propuesta por dos de nosotros (Golub y Fendlebury) daría verdadera refrigeración, al emplear helio líquido como medio refrigerante. A temperaturas inferiores a unos dos grados Kelvin, el helio se convierte en superfluido: pierde su viscosidad y exhibe propiedades mecánico cuánticas a escala macroscópica. Un superfluido posee pocos estados de movimiento, cada una de las cuales ostenta una energía y un momento bien determinados. El intercambio de energía entre un neutrón y el superfluido puede ocurrir sólo si las cantidades de energía y de momento intercambiadas son respectivamente iguales a las diferencias de energía y de momento entre dos de estos estados superfluidos. Un neutrón con una energía en torno a un milielectronvolt (o una longitud de onda de 10 angstrom) puede transformarse en neutrón perteneciente al intervalo de energías ultrafrías por colisión con el superfluido. La energía suministrada por el neutrón crea una excitación mecánico cuántica en el helio que recibe el nombre de fonón, el cual se desintegrará posteriormente. Cuando el neutrón ha pasado a ser ultrafrío puede ganar energía del helio a través del mecanismo opuesto, en el que un fonón cede la energía a un neutrón, quedando éste con una energía de alrededor de milielectronvolt. Si la temperatura se mantiene suficientemente baja, por debajo de un grado Kelvin, los fonones son raros y la colisión a estados de mayor energía resulta poco probable.

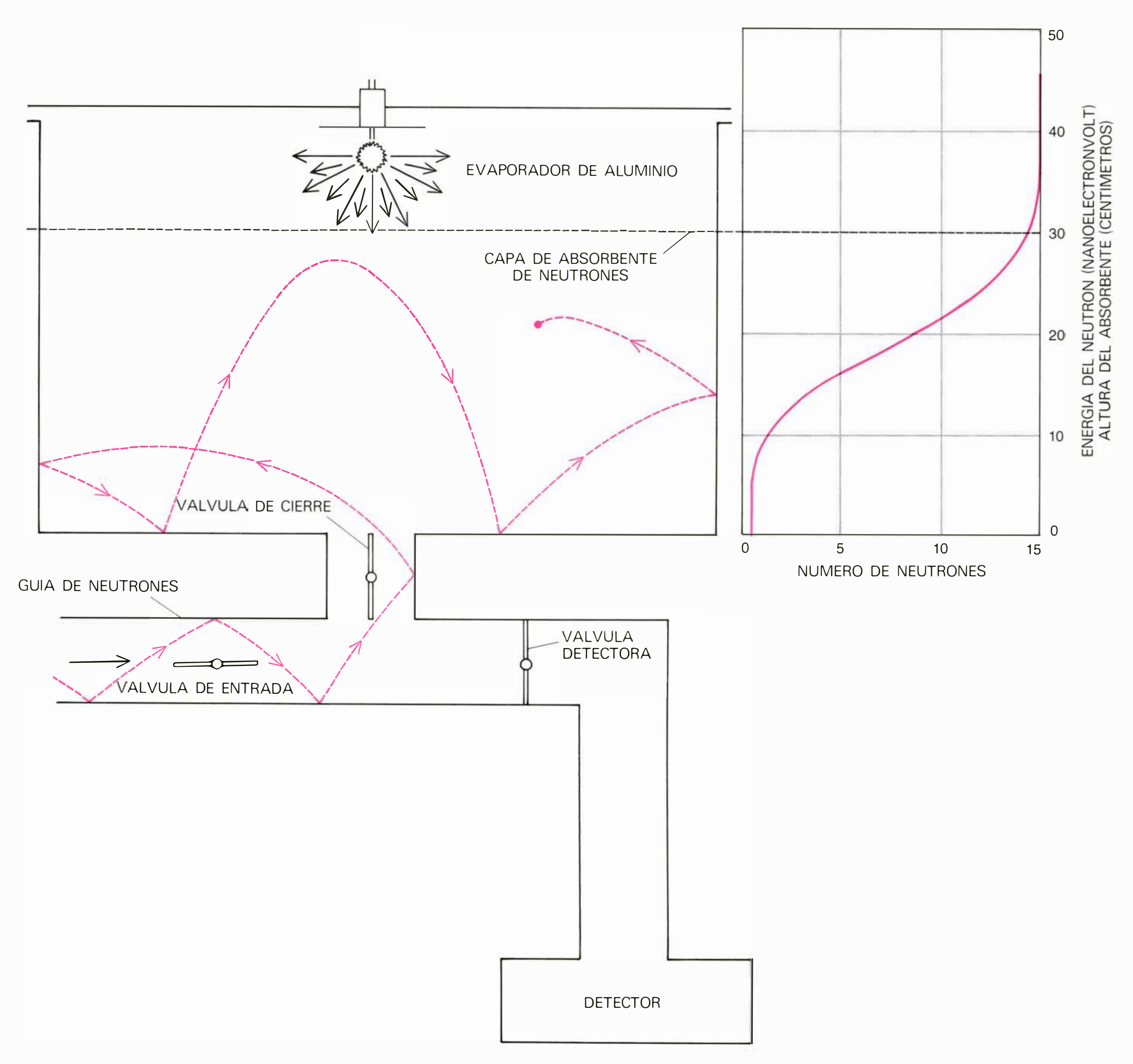
Los neutrones no sólo suministran energía al superfluido sino que también le ceden calor. La distribución final de energía queda confinada en un intervalo menor. Si el superfluido se coloca en un flujo de neutrones de 10 angstrom, el calor pasa de forma continua al helio. Si extraemos luego del helio el calor a través de un sistema de refrigeración externo, se crean neutrones ultrafríos a un ritmo constante. No hace mucho pudimos mostrar la posibilidad de crear neutrones ultrafríos por este método en el Instituto Laue-Langevin.

Si introducimos el helio en una vasija apropiada, los neutrones ultrafríos que se creen en el superfluido quedarán atrapados en las paredes. Se espera poder acumular neutrones a densidades mucho más altas que las obtenidas por

otros métodos. Las pérdidas en el helio son despreciables. Ya hemos visto que las colisiones inelásticas a energías más altas son improbables, y si el helio está constituido exclusivamente por el isótopo común helio 4, la absorción es imposible. Un núcleo de helio 4 no puede absorber un neutrón. Además, el enfriamiento del aparato debería también reducir considerablemente las pérdidas en las paredes causadas por la colisión inelástica con las impurezas. Otro mecanismo de pérdida importante puede evitarse por completo: el agujero de entrada de la vasija, por donde los neutrones podrían escapar. No hay necesidad de una abertura, ya que las paredes pueden hacerse transparentes a los neutrones incidentes de 10 angstrom y opacas a los de longitud de onda mayor creados dentro de la vasija.

Cuando un neutrón alcanza el intervalo de energías ultrafrías en esta fuente de helio, hay gran posibilidad de que permanezca allí hasta que se desintegre. En muchos experimentos acometidos con sistemas convencionales de extracción y almacenaje, la densidad de neutrones ultrafríos es menor de uno por centímetro cúbico. Estimamos que con una fuente superfluida colocada fuera de un reactor al final de un tubo guía podría aumentarse la densidad de neutrones ultrafríos en un factor de 10 a 100. Tal fuente se halla en fase de construcción en el Instituto Laue-Langevin.

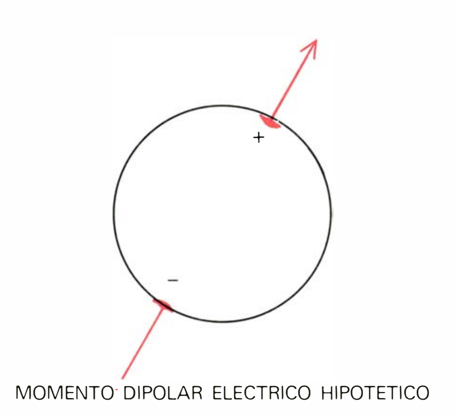
La botella esférica magnética que están preparando Paul y Trinks en Bonn operará con la fuente superfluida. La

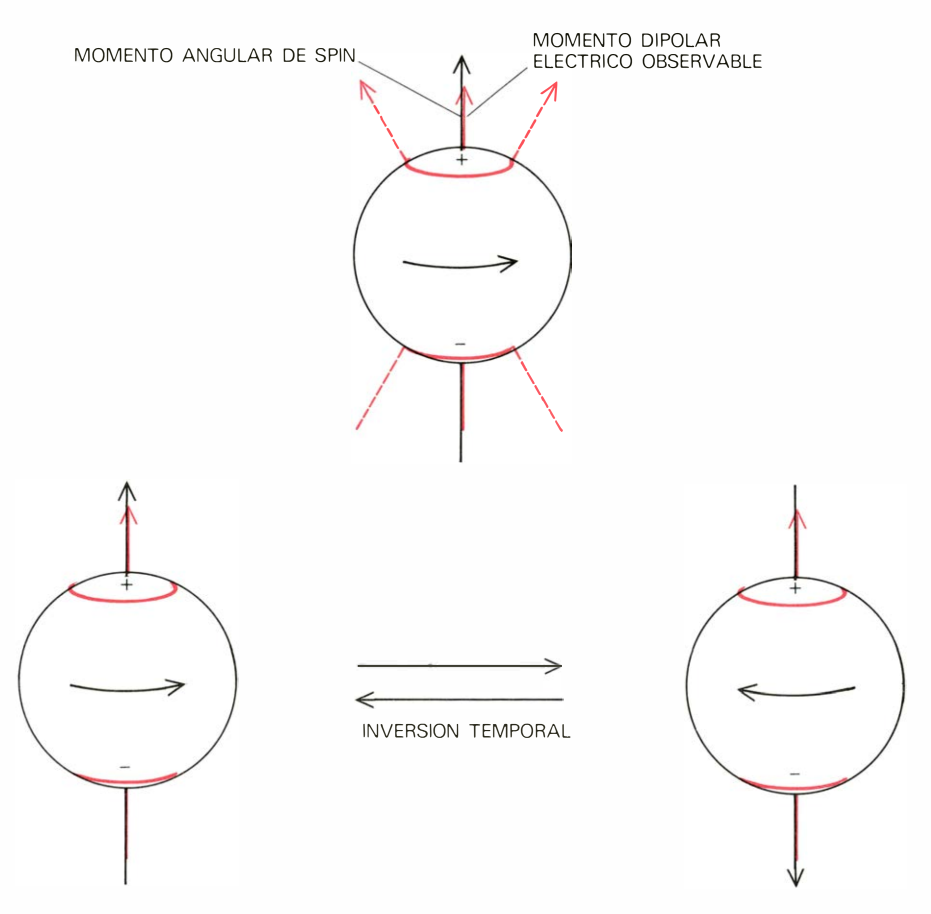


UNA SUPERFICIE CONTINUAMENTE RENOVADA y neutrones moviéndose a un poco más de un metro por segundo se emplearon en otro experimento por Strelkov y sus colaboradores. Se permitía la entrada de neutrones a través de una válvula, se sellaba la cámara y se colocaba a una altura predeterminada una capa de material absorbente de neutrones. Los neutrones que conseguían alcanzar el absorbente quedaban atrapados por éste, de suerte que sólo los que se movían con excesiva lentitud para llegar a la altura del absorbente podían sobrevivir. Después de un cierto período, se contaban abriendo las válvulas a un detector. El espectro de energía de los neutrones (*derecha*) se determinaba midiendo el número de neutrones que quedaban en la vasija como una función de la altura del absorbente. En un esfuerzo por reducir la contaminación de la superficie se evaporaba aluminio sobre las paredes de la cámara, pero no se pudo detectar ningún efecto del material depositado sobre el tiempo de almacenamiento. La hipótesis de las impurezas continúa siendo, hoy en día, la más aceptable, porque no se ha medido el contenido en hidrógeno de las capas del aluminio evaporado.

esfera se llenará con helio líquido en presencia del campo magnético; y se expondrá a un haz de neutrones de 10 angstrom. Cuando se haya acumulado en su interior un número suficiente de neutrones ultrafríos, se extraerá el helio, pero los neutrones quedarán atrapados por el gradiente de campo magnético. El único mecanismo importante de pérdida deberá ser la desintegración beta de los propios neutrones.

Una de las razones para mejorar el tiempo de almacenaje de los neutrones ultrafríos es poder medir con precisión la vida media del neutrón. Como ya hemos dicho, un neutrón fuera de un núcleo se desintegra dando un protón, un electrón y un neutrino. Este proceso de la desintegración beta viene mediado por una de las cuatro fuerzas básicas de la naturaleza, la llamada fuerza débil. La desintegración es un proceso estadístico y no puede predecirse cuánto tiempo sobrevivirá un neutrón determinado, aunque sí medir una duración media.





MOMENTO DIPOLAR ELECTRICO del neutrón. Si fuera mayor que cero, ello significaría una violación de la simetría de inversión temporal. Un momento dipolar eléctrico es una separación de cargas eléctricas, dando a la partícula un extremo positivo y otro negativo. El momento puede representarse como un vector, que se simboliza aquí dirigido hacia el polo positivo. El momento angular de spin del neutrón es también un vector. Ambos vectores deben hallarse alineados según el mismo eje, o al menos la única componente del momento dipolar que puede ser observada es la que está alineada con el vector del spin. La simetría bajo la inversión temporal exige que el estado de una partícula permanezca sin cambiar si se modifica el sentido de todos los movimientos. En algún estado inicial los dos vectores pueden ser paralelos. Si invirtiéramos el tiempo, se invertiría la dirección del vector de spin, pero el vector momento dipolar eléctrico no quedaría afectado. Los dos estados no serían pues equivalentes.

El valor actualmente aceptado de la vida media de un neutrón libre es de 918 segundos, es decir, 15 minutos y 18 segundos. La precisión de la medida tiene un margen de 14 segundos en más o en menos que representan un error probable de 1,5 por ciento. Se conoce la vida media de muchas partículas menos familiares con una precisión mucho mayor. Por ejemplo, el muón, que también se desintegra bajo la acción de la fuerza débil, tiene una vida media de sólo dos millonésimas de segundo, pero se ha medido con una precisión de 0,003 por ciento. Una medición más exacta de la vida media del neutrón permitiría comprender mejor en qué consiste la fuerza débil.

Una forma directa de medir una vida media es comenzar con un número conocido de partículas y contabilizar el número de las que quedan en función del tiempo. El resultado sería una curva exponencial a partir de la cual se deduciría fácilmente la vida media. La principal dificultad para determinar la vida media del neutrón por este procedimiento estriba en que se trata de una partícula sin carga eléctrica y no puede detectarse con muy buena resolución. La única forma práctica de detectar un neutrón lento es dejar que sea absorbido por un núcleo, el cual elimina posteriormente el exceso de energía en forma de partículas cargadas energéticas. La partículas cargadas son las realmente detectadas. Un núcleo que suele utilizarse en detectores de neutrones es el de helio 3, el isótopo menos común del helio, que absorbe neutrones con gran avidez.

Toda medición normal de la vida media del neutrón, llevada a cabo con un haz de neutrones no confinados, exige dos detectores: un detector de neutrones, que cuenta los neutrones del haz, y un detector sensible a los electrones o a los protones, que, a su vez, cuenta los neutrones que se han desintegrado. El cálculo de la vida media depende de la eficiencia de ambos contadores, los cuales poseen una precisión muy limitada.

Un aparato adecuado para almacenar neutrones simplificaría considerablemente el procedimiento, de suerte que sólo sería necesario un detector. Midiendo el número de neutrones presentes en la vasija de almacenamiento como una función del tiempo de almacenaje,

APANTALLAMIENTO MAGNETICO

CAMPO MAGNETICO ESTACIONARIO

CAMPO MAGNETICO OSCILANTE

VALVULA

FUENTE DE ALTA TENSION

CAMPO ELECTRICO

PLACAS CONDUCTORAS

AISLADOR

TUBO GUIA DE NEUTRONES

LAMINA POLARIZANTE

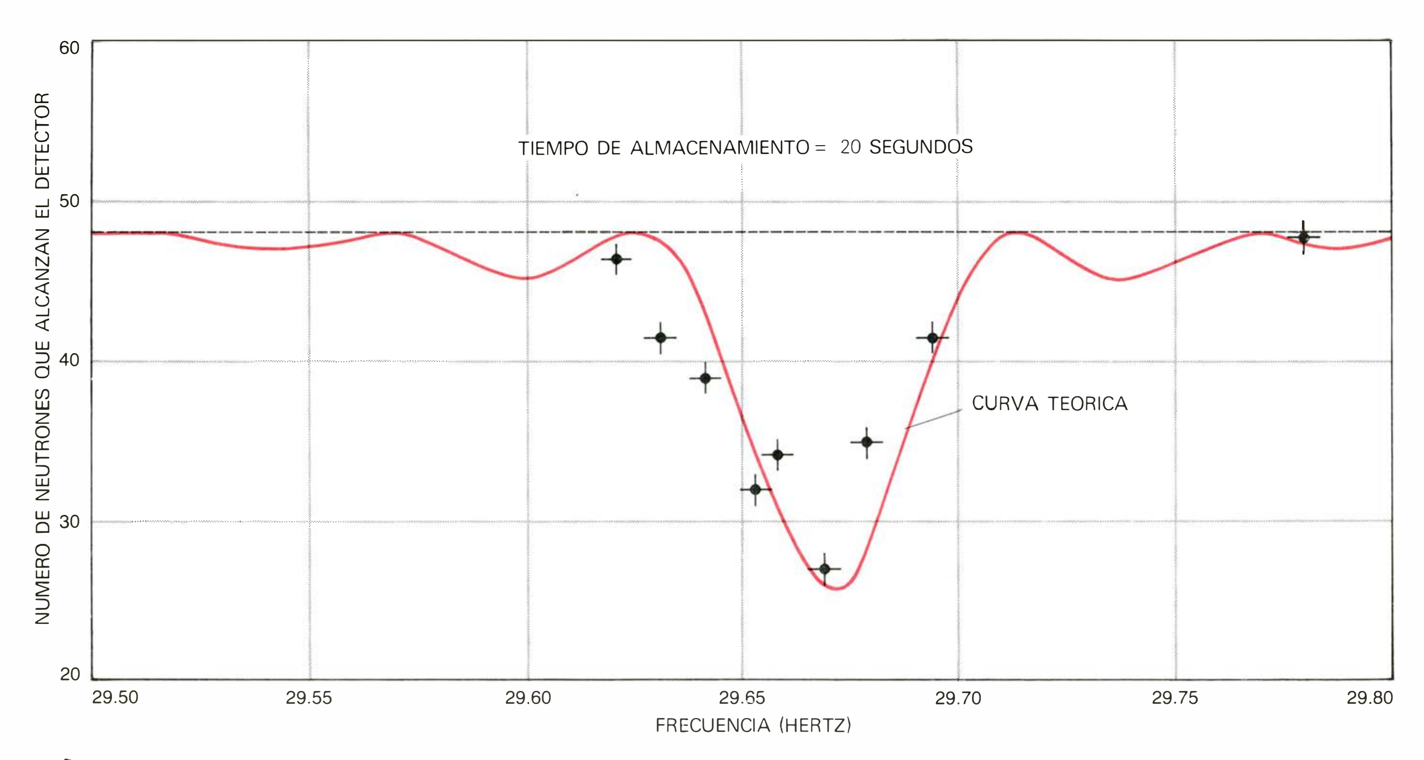
VALVULA

FUENTE DE NEUTRONES ULTRAFRIOS

DISPOSITIVO PARA UN CAMPO MAGNETICO ESTACIONARIO

DETECTOR

BUSQUEDA EXPERIMENTAL de un momento dipolar eléctrico del neutrón. Se pretende con ello detectar un pequeño cambio en la energía del neutrón cuando está sometido a un campo eléctrico muy intenso. El campo eléctrico está generado por un voltaje alto aplicado entre dos placas conductoras; la cámara está sumergida también en un campo magnético estacionario. Los neutrones entran en la cámara a través de una lámina polarizadora que deja pasar sólo aquellos neutrones cuyo momento magnético es paralelo al campo magnético estacionario. Un campo magnético oscilante invierte luego el spin y el momento magnético del neutrón, por lo que no pueden atravesar de nuevo la lámina para alcanzar el detector. La frecuencia de oscilación necesaria para inducir este cambio de orientación del spin depende de la energía de los neutrones, condicionada por el momento eléctrico dipolar. Cambios en la magnitud del campo eléctrico desplazarían la frecuencia de resonancia si el neutrón presentase un momento dipolar.



SEÑAL DE RESONANCIA registrada en el Instituto Laue-Langevin; constituye un indicador sensible a la energía del neutrón. El acusado pozo de la curva obedece al cambio de orientación, de 180 grados, de los spines y momentos magnéticos de los neutrones cuando un campo magnético oscilante se sintoniza a la frecuencia de resonancia. Los neutrones que han girado su spin se muestran incapaces de atravesar la lámina polarizadora y, por tanto, no pueden alcanzar el detector. Un momento dipolar eléctrico se detectaría como un desplazamiento en la posición del mínimo. La precisión con que se puede medir cualquier frecuencia es proporcional al tiempo de duración de la medida. La posibilidad de almacenar neutrones ultrafríos durante intervalos de tiempo comparativamente largos haría posible establecer una nueva cota superior para el momento dipolar del neutrón.

podríamos determinar el ritmo de pérdida. Si pudieran despreciarse las pérdidas debidas a neutrones que se escapen, entonces el ritmo medido quedaría determinado totalmente por la vida media de la desintegración beta.

El grupo de Bonn ha realizado mediciones preliminares de la vida media con neutrones ultrafríos. Pero debe conseguirse una precisión todavía mayor. Es más, cabe esperar que pronto dispondremos de una medición definitiva. En la botella magnética esférica, que se empleará para la determinación de la vida media en el Instituto Laue-Langevin, se espera que sólo las indeterminaciones estadísticas (en oposición a las experimentales) limiten la precisión de la medición. Las indeterminaciones estadísticas dependen sólo del número de neutrones disponibles (y de la paciencia del experimentador.)

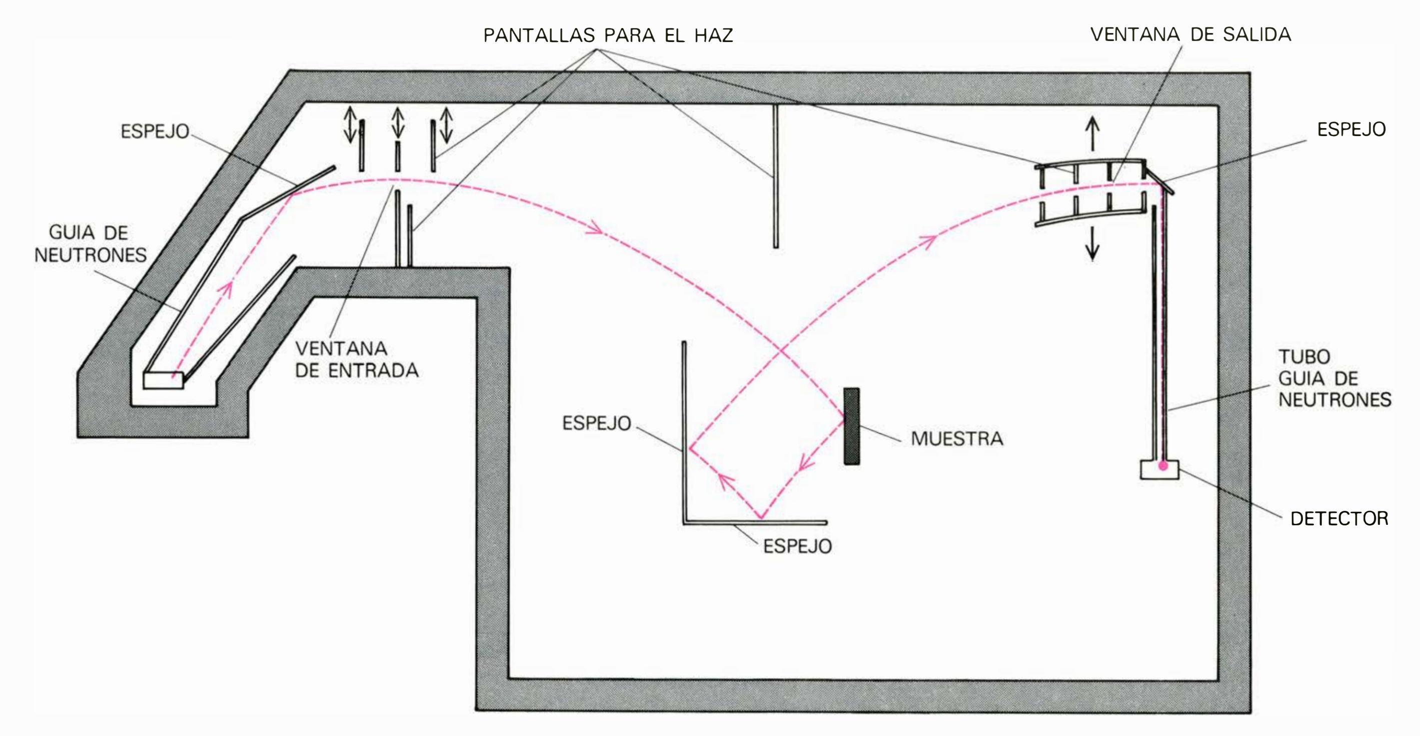
De la obtención de tiempos de almacenamiento mayores saldría ganando también la búsqueda de un momento dipolar eléctrica del neutrón. La carga eléctrica total del neutrón es, por supuesto, cero, pero podría suceder que hubiera cargas en el "interior" de la partícula y que su carácter neutro se deba sólo a que las cargas componentes se cancelen unas a otras, es decir, que entre todas sumen cero. (En efecto, la teoría actual del neutrón, que considera que el neutrón está compuesto de partículas más pequeñas llamadas quarks, predice precisamente una estructura como ésta.) Si las cargas positivas y negativas están desplazadas de modo permanente, o si son móviles pero tienden a pasar mucha parte del tiempo en regiones distintas, entonces el neutrón tendrá un momento dipolar eléctrico: poseerá un extremo positivo y otro negativo. La magnitud del momento puede ser muy pequeña, pero las consecuencias del mismo no serían triviales. El descubrimiento de un momento dipolar eléctrico en el neutrón revelaría la violación de un principio físico que se conoce con el nombre de simetría bajo la inversión temporal.

En física la palabra simetría tiene un sentido más amplio que en el lenguaje ordinario. Un principio de simetría establece que un objeto o un proceso permanece inalterado cuando se hace una determinada transformación. Por ejemplo, el principio de simetría con respecto a las traslaciones espaciales establece que una experiencia realizada en Alma-Ata dará el mismo resultado que cuando se haga en Grenoble, que está a 8000 kilómetros.

Dos importantes simetrías en física de partículas elementales son las de paridad e inversión temporal. El principio de simetría con respecto a la paridad implica que ningún proceso físico puede distinguir la derecha de la izquierda. Si la paridad fuera una simetría válida de la naturaleza, cualquier suceso permitido por las leyes de la física también sería permitido si derecha e izquierda se intercambiaran, como en un espejo. En 1957 C. S. Wu y sus colegas, de la Universidad de Columbia, demostraron que la paridad no es una simetría válida en todas las situaciones. Probaron que se viola en la desintegración beta de los núcleos de cobalto 60; y se sabe ahora que la paridad es violada en todas las desintegraciones beta, incluida la del neutrón. (En algunas otras interacciones, sin embargo, la paridad se conserva estrictamente.)

La simetría de inversión temporal establece que cualquier proceso permitido en la naturaleza se permitiría también si cambiáramos el sentido de todos los movimientos, o en otras palabras, si el tiempo transcurriera hacia atrás. A escala macroscópica, es inmediato que el principio se viola en vez de cumplirse: un jarro se rompe, pero no vuelve a recomponerse. Sin embargo, este proceso inverso está prohibido únicamente por su improbabilidad; no violaría ninguna de las leyes físicas conocidas. A nivel de las partículas subatómicas, muchos procesos parecen ser totalmente reversibles. Del mismo modo que un neutrón se desintegra en un protón, un electrón y un neutrino, también un protón, un electrón y un neutrino pueden unirse para formar un neutrón.

En física de partículas se ha descubierto una violación de la simetría de

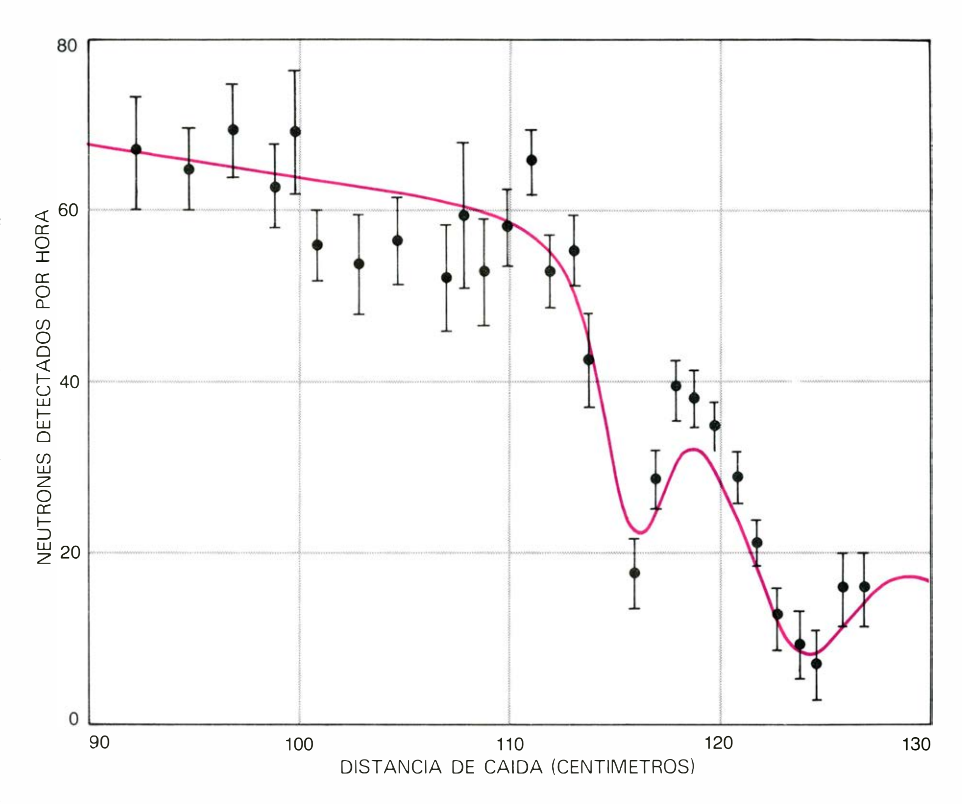


ESPECTROMETRO DE NEUTRONES construido por Albert Steyrel y sus colaboradores en la Universidad Técnica de Munich. Emplea el campo gravitacional terrestre para analizar la energía de los neutrones ultrafríos que han rebotado en un determinado material de muestra. Los neutrones entran en el aparato moviéndose horizontalmente y se aceleran mientras caen hacia la muestra. Después de reflejarse en ella y en un par de espejos, los neutrones ascienden tanto como les sea posible son recogidos por una rendija de salida ajustable. La altura hasta la que han llegado tras el rebote es sensible a variaciones mínimas de la energía de los neutrones. Este aparato sirve para estudiar películas delgadas y superficies de sólidos.

inversión temporal. Está relacionada con la desintegración del mesón neutro *K*, y se encontró en una serie de experimentos que empezaron en 1964 con el trabajo de James H. Christenson, James W. Cronin, Val L. Fitch y René Turlay, de la Universidad de Princeton. No se ha encontrado ningún otro ejemplo de violación de inversión temporal. Más aún, no se sabe todavía qué fuerza fundamental de la naturaleza es la responsable de dicha violación. La fuerza débil es el candidato más probable, ya que es responsable de las desintegraciones beta y, por tanto, de la violación observada de la paridad. Las otras posibilidades son las interacciones electromagnéticas y una nueva e hipotética fuerza, llamada fuerza superdébil.

La observación de un momento dipolar eléctrico de cualquier partícula elemental evidenciaría otra violación de la simetría de la inversión temporal. La razón de esto es la siguiente: si la partícula tiene un momento dipolar eléctrico, el eje que conecta los polos eléctricos debe estar alineado con el eje de rotación de spin. Si el momento dipolar fuera perpendicular al eje de spin, de suerte que los polos eléctricos estuvieran en el ecuador, entonces el giro de la partícula daría un momento medio nulo, igual que un disco blanco y negro aparece gris cuando gira rápidamente. Si el dipolo tuviera alguna orientación intermedia entre la paralela y la perpendicular, sólo sería observable la componente del momento dipolar alineada con el eje de spin.

Tanto el spin como el momento dipolar eléctrico pueden representarse como vectores o flechas. La dirección del vector de spin se acostumbra definir mediante la regla de la mano derecha: si los dedos de la mano derecha rodean la partícula en su dirección de giro, el pulgar da entonces la dirección del vector. Adoptaremos una convención similar para el vector dipolar eléctrico: pasa a través de los dos polos eléctricos y apunta en la dirección de la carga positiva. Supongamos que, en algún estado inicial, el vector de spin y el vector dipolar eléctrico están apuntando en la misma dirección, digamos hacia arriba. Bajo inversión temporal, la partícula debe girar en sentido opuesto y el vector de spin ha de señalar hacia abajo. La inversión temporal no produce ningún efecto sobre el momento dipolar eléctrico, como tampoco afecta la carga de la partícula. Luego, en un estado, los vectores son paralelos y, en el estado obtenido por inversión temporal, son antiparalelos; los dos estados no son equivalentes y se viola la simetría bajo



ESPECTRO DE NEUTRONES reflejado a partir de una fina película de oro aplicada a uno de los espejos en el espectrómetro de neutrones. Como puede apreciarse, el espectro muestra un modelo de interferencia, una señal inequívoca de la naturaleza ondulatoria del neutrón. La altura desde donde caen los neutrones antes de golpear la película varía a lo largo de una gama de 40 centímetros, lo que origina una variación asociada de la energía de los neutrones detectados en la ventana de salida, corresponden a las energías en que las ondas de neutrones se interfieren constructivamente con la película de oro. Según se observa en dos valles de la curva, las ondas se interfieren destructivamente.

inversión temporal. Hay una total diferencia con lo que sucede con el momento dipolar magnético, que depende de la dirección del spin: si hacemos una inversión temporal, el dipolo magnético cambia de dirección al propio tiempo que el vector de spin.

La búsqueda de un momento eléctrico dipolar se ha centrado principalmente en el neutrón, porque los experimentos necesarios se acometen mejor con neutrones que con otras partículas. Los experimentos con neutrones de alta energía han establecido ya un límite máximo sobre el momento dipolar eléctrico del neutrón: si existe, debe ser inferior a $3 \times 10^{-24} e^{-}$ centímetros, donde e es la carga del electrón. En 1968 Shapiro, en Dubna, hizo notar que podía obtenerse un límite todavía menor mediante experimentos con neutrones ultrafríos.

La técnica experimental básica consiste en exponer los neutrones a un intenso campo eléctrico; si el neutrón tiene un momento dipolar eléctrico, su energía dependerá en una pequeña proporción de la orientación del dipolo eléctrico con respecto al campo eléctrico aplicado. Si el neutrón cambia de orientación 180 grados, la energía deberá cambiar en una cantidad proporcional a la intensidad del dipolo y a la intensidad del campo exterior.

En la práctica hay que recurrir a un método indirecto para observar este pequeño cambio de energía. Se establece una diferencia de energía mucho más alta mediante un campo magnético constante y fuerte, que actúa sobre el momento dipolar magnético (no sobre el eléctrico) del neutrón. Se obliga luego a que los neutrones cambien su orientación en 180 grados mediante un campo magnético oscilante. La frecuencia exigida para inducir este salto del spin depende, de forma muy precisa, de la energía exacta de los neutrones y, por tanto, viene determinada en parte por la interacción del momento dipolar eléctrico con el campo eléctrico aplicado. Si se cambia la magnitud del campo eléctrico, y si existe el momento dipolar eléctrico del neutrón, la frecuencia de resonancia habrá de cambiar. Si es nulo el momento dipolar, el campo eléctrico no afectará la frecuencia necesaria para cambiar en 180 grados la dirección del spin.

Cualquier medición de una frecuencia es incierta en una oscilación, pero la importancia de esta incertidumbre disminuye a medida que el número de ciclos contados aumenta. Los neutrones ultra-

fríos dan un método muy preciso para medir el momento dipolar eléctrico del neutrón, pues permiten un tiempo largo de observación y, por tanto, una determinación cabal de la frecuencia de resonancia. V. M. Lobashov y sus colegas, del Instituto de Física Nuclear de Leningrado, acaban de enunciar los primeros resultados de un experimento sobre el momento dipolar eléctrico del neutrón al emplear neutrones ultrafríos. En su trabajo, no se almacenaban los neutrones sino que pasaban lentamente por los campos eléctrico y magnético aplicados con un tiempo de contaje de unos cuatro segundos. Los resultados son provisionales, y aunque ellos no mejoran el límite sobre el momento dipolar, demuestran que esta técnica es capaz de una precisión mayor que otros métodos. En un experimento que se halla ahora en fase de preparación en el Instituto Laue-Langevin, nosotros emplearemos neutrones confinados y esperamos obtener una nueva cota superior.

Tradicionalmente, los neutrones térmicos se vienen aplicando para indagar las estructuras y movimientos atómicos y moleculares, sobre todo en sólidos. Los neutrones complentan los rayos X y otras formas de radiación electromagnética que se utilizan en estos estudios: los rayos X son sensibles a los electrones en los sólidos, mientras que los neutrones interaccionan, primordialmente, con los núcleos atómicos. Los neutrones ultrafríos pueden usarse de forma análoga, pero su baja energía y su larga longitud de onda los hace adecuados para examinar los materiales a una escala algo mayor.

Steyerl ha construido un nuevo espectrómetro de neutrones ultrafríos. En un espectrómetro ordinario, donde los proyectiles son partículas cargadas, se analiza la energía de las partículas mediante un imán que doble sus trayectorias. En el espectrómetro de neutrones ultrafríos, el papel de analizador lo desempeña el campo gravitacional de la Tierra. Los neutrones entran en el espectrómetro moviéndose horizontalmente y sufren luego una aceleración al caer a lo largo de una distancia fija hasta la muestra. Después de la colisión, los neutrones se elevan todo lo que pueden contra la fuerza gravitacional y se recogen en una rendija de salida. Si la interacción con la muestra es perfectamente elástica, se elevarán al mismo nivel desde donde han caído. Cualquier intercambio de energía con la muestra se detectará como un cambio en la altura máxima posible de los neutrones después de la colisión.

Como la energía total de los neutrones es muy pequeña (unos 10^{-7} electronvolt), el aparato es extraordinariamente sensible a pequeños cambios de energía causados por las interacciones con la muestra. El espectrómetro puede darnos un conocimiento más preciso de los movimientos de los átomos en los sólidos y de las fuerzas entre ellos. Puesto que los neutrones ultrafríos penetran sólo unos 100 angstrom, el espectrómetro debería ser adecuado para estudiar películas delgadas y superficies de sólidos. Steyerl ha demostrado la interferencia de neutrones reflejados por películas de un espesor de 1000 angstrom.

La difracción de neutrones (de electrones e incluso de átomos) ha sido comprobada muchas veces y es la mejor prueba de su naturaleza ondulatoria. La difracción se produce generalmente haciendo pasar las partículas a través de la red ordenada de los átomos en un cristal, donde la distancia entre átomos es del orden de un angstrom. Steyerl ha demostrado la difracción de neutrones ultrafríos por una red de difracción, idéntica a las usadas para la luz. El experimento es una manifestación directa de un efecto mecánico cuántico: demuestra que la onda asociada al neutrón se halla realmente distribuida sobre una distancia de unos 1000 angstrom. Tales experimentos de difracción permitirán una comparación directa de una longitud de onda óptica con una longitud de onda de neutrones, y quizá conduzcan a una nueva determinación de una de las constantes fundamentales de la naturaleza: la constante de Planck.

Steyerl (con G. Shütz) ha iniciado un nuevo proyecto que consiste en la construcción de un microscopio de neutrones. Los neutrones serán enfocados mediante una lente de Fresnel, que funciona usando las interferencias y que no podría construirse para neutrones de longitudes de onda más cortas. De nuevo, el aparato podría complementar uno ya existente, el microscopio electrónico. Algunos materiales que son indistinguibles o transparentes a los electrones revelan contrastes acusados cuando se les ilumina con neutrones.

El desarrollo de estos instrumentos y el progreso reciente de los experimentos sobre la vida media del neutrón y su momento dipolar eléctrico sugieren que una primera fase en el estudio de los neutrones ultrafríos puede estar terminada. La década transcurrida desde que se detectaron por primera vez se ha empleado en aprender cómo extraerlos, almacenarlos y manipularlos. Se puede ahora hacer buen uso de los mismos para estudios sobre el neutrón y otros sistemas de partículas.

La historia del Atlántico

El océano tiene una edad de 165 millones de años. Su crecimiento y la topografía de su fondo se explican gracias a los testigos de sondeo, las mediciones del flujo de calor y del magnetismo y al desarrollo de la teoría de la tectónica de placas

John G. Sclater y Christopher Tapscott

Hace 165 millones de años, el océano Atlántico no existía. Los continentes que en la actualidad lo bordean se encontraban entonces unidos, cual piezas de un rompecabezas, formando una inmensa extensión de tierra. Por esa época, el magma fundido que ascendía desde el interior del planeta formó los continentes al producir la separación de las tierras emergidas. Como consecuencia de ello, los continentes empezaron a separarse a una velocidad de varios centímetros por año; y así nació el Atlántico. Durante las dos últimas décadas los geofísicos han cartografiado la topografía del fondo del Atlántico, para lo cual ha sido literalmente bombardeado con ondas sonoras de las que se han estudiado sus ecos. La estructura vertical del fondo oceánico se ha determinado a partir de muestras cilíndricas de sedimentos, recogidas por buques especializados en sondeos en mares profundos, habiéndose medido también las anomalías de la magnitud y dirección del campo magnético terrestre existente en esos mismos fondos. La tectónica de placas y los modelos teóricos actuales sobre el flujo de calor en el fondo oceánico pueden explicar los rasgos topográficos, las principales características de la sedimentación y las anomalías magnéticas. La teoría y los datos obtenidos han permitido a los científicos que estudian la tierra reconstruir la historia del Atlántico y, en particular, la historia de su batimetría o profundidad.

El rasgo topográfico más sobresaliente del Atlántico es la dorsal centro-oceánica, enorme cordillera montañosa sumergida, cuya escala sobrepasa a los Alpes y al Himalaya juntos, que recorre la parte central del océano desde Islandia, al norte, hasta la isla de Bouvet (a 1800 kilómetros de distancia de la costa antártica), al sur. La cresta de la dorsal marca aproximadamente la parte media del Atlántico y, a excepción de determinadas zonas aisladas, cerca de Islandia y cerca de las Azores, el fondo oceánico en esa zona de la cresta tiene una profundidad media de 2500 metros. La profundidad del fondo aumenta gradual y simétricamente con la distancia a la cresta, alcanzando una profundidad máxima, comprendida entre 5000 y 6000 metros, antes de levantarse abruptamente en las plataformas continentales.

Esta simetría norte-sur, tan simple, del fondo del Atlántico queda interrumpida por dorsales transversales aisladas y por zonas de fractura, es decir, por cicatrices profundas del fondo oceánico, que cortan a la cresta de la dorsal central y siguen una dirección aproximadamente perpendicular a ella. Las dorsales este-oeste se levantan unos 2000 metros por encima del fondo oceánico circundante. A diferencia de la dorsal central, donde se localizan los epicentros de la mayoría de los terremotos superficiales que se producen en el Atlántico, las dorsales este-oeste son sísmicamente inactivas. Entre estas últimas, las dorsales de Walvis y de Río Grande son las más importantes, abarcando entre las dos la parte central del Atlántico sur. Los cañones producidos por las zonas de fractura (que pueden llegar a tener hasta 500 kilómetros de longitud, 25 kilómetros de anchura y 3 kilómetros de profundidad) desplazan el eje de la dorsal centro-oceánica de forma intermitente a lo largo de toda su longitud.

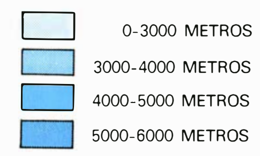
En la cresta de la dorsal centro-oceánica el fondo está constituido, fundamentalmente, por rocas volcánicas duras y resistentes. Al aumentar la distancia a la cresta, estas rocas aparecen recubiertas por un espesor creciente de sedimento blando, constituido por una combinación de arcilla roja (material detrítico ligero procedente de los continentes) y barro calcáreo (carbonato cálcico procedente de restos esqueléticos de microorganismos). Cerca de la cresta de la dorsal prodomina el barro calcáreo. Entre 4,5 y 5 kilómetros por debajo de la superficie oceánica se encuentra el nivel de compensación del carbonato cálcico que es la profundidad por debajo de la cual la lluvia continua de organismos pelágicos, en forma de sedimentos calcáreos, queda disuelta en vez de acumularse debido a que el agua no está saturada de carbonato cálcico. Esto significa que, en las aguas profundas alejadas de la cresta de la dorsal centro-oceánica, los sedimentos de carbonato cálcico se disuelven, y así la arcilla roja pasa a constituir el material predominante sobre el fondo.

En la vecindad de los continentes, los sedimentos se recubren de capas de material detrítico procedente de las plataformas continentales. En esas zonas el fondo oceánico es extremadamente uniforme, formado por espesas capas de sedimentos que constituyen las suaves llanuras abisales. La acumulación máxima de sedimentos se encuentra al pie de los continentes. Los sedimentos son depositados por los ríos y están constituidos principalmente por finas partículas de arcilla y arena mezcladas con barro calcáreo y materiales de origen orgánico.

Desde 1968, el buque oceanográfico *Glomar Challenger,* equipado para realizar sondeos en mar profundo y construido por un consorcio de instituciones oceanográficas americanas, ha reunido columnas de sedimentos del fondo oceánico procedentes de centenares de sitios. Uno de los resultados más sorprendentes fue descubrir la existencia de sedimentos formados por carbonato cálcico en muestras obtenidas cerca de la plataforma continental. Delgadas capas de sedimentos de este tipo, encontradas directamente encima de rocas volcánicas duras, fueron cubiertas, a su vez, por hasta un kilómetro de espesor de arcilla roja y otros materiales sedimentarios resultantes de la erosión continental. El carbonato cálcico se encontraba entre seis y siete kilómetros por debajo de la superficie oceánica. Puesto que el nivel de compensación queda muy por encima, la presencia de carbonato cálcico resultaba desconcertante. El enigma se deshizo cuando pudo reconstruirse la historia batimétrica del Atlántico.

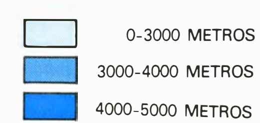
DORSAL DE RIO GRANDE

DORSAL DE WALVIS



MAPA BATIMETRICO DEL ATLANTICO donde se observa que el fondo oceánico aumenta de profundidad, simétricamente, al crecer la distancia a la dorsal centro-oceánica, la gran cordillera sumergida que desciende por la parte media del océano y que va desde Islandia, al norte, hasta la isla de Bouvet, al sur, situada a 1800 kilómetros de la costa antártica. El fondo oceánico, que tiene una profundidad media de 2500 metros de la cresta de la dorsal, alcanza una profundidad máxima entre 5000 y 6000 m antes de levantarse abruptamente en las plataformas. La simetría norte-sur del fondo queda interrumpida por dorsales asísmicas y zonas de fractura estrechas que, a modo de profundas cicatrices, los accidentan. La dorsal de Walvis y la de Río Grande son las asísmicas más prominentes.

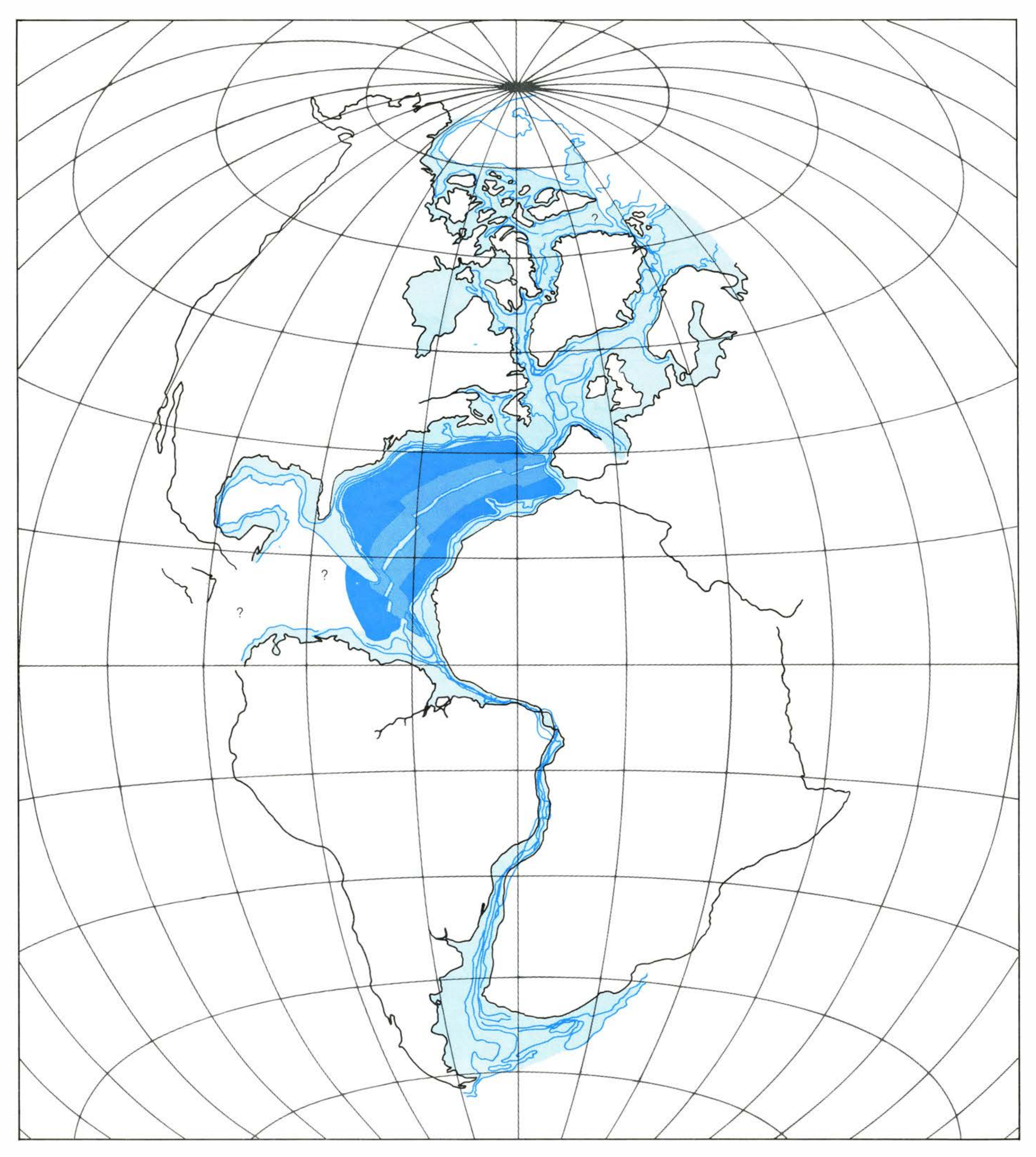
Muchos de los fenómenos geofísicos a gran escala que se producen en el fondo oceánico pueden explicarse gracias a la teoría de la tectónica de placas, de acuerdo con la cual se considera la superficie de la tierra formada por un conjunto de placas rígidas que forman la litosfera. Las placas varían de tamaño y así tenemos desde las seis principales, como la placa que constituye prácticamente todo el Pacífico, hasta otras más pequeñas, como la que coincide casi con la actual Turquía. La litosfera mide unos 100 kilómetros de espesor, y comprende la corteza y el manto superior de la tierra. Flota sobre la astenosfera, que es una capa móvil e íntegramente fluida del manto con un espesor de varios centenares de kilómetros. Las placas litosféricas flotan sobre la astenosfera de forma parecida a como la cera sólida se sostiene en un recipiente con cera fundida. Las placas tienen siempre movimientos relativos entre sí, por lo que in-



MAPA DE LOS CONTINENTES hace 165 millones de años (*izquierda*). América del Norte, América del Sur, Groenlandia, Europa y Africa, que habían formado una enorme extensión de tierra, empezaron a separarse a medida que se fue produciendo entre ellos la ascensión de magma fundido, procedente de las zonas profundas. La creación de nueva corteza, gracias al flujo de magma hacia el exterior, separó los continentes septentrionales de Africa y América del Sur, abriendo el naciente Atlántico septentrional y Caribe. Hace 125 millones de años (*derecha*) el Atlántico Norte había alcanzado profun-

terfieren mutuamente en sus zonas de contacto.

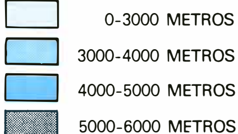
Las placas pueden interferirse de tres maneras: por divergencia, es decir, separándose (tal como ocurre en los lados opuestos de la dorsal centro-atlántica); por convergencia, es decir, las placas pueden colisionar; por último, por deslizamiento, moviéndose lateralmente entre sí. En la dorsal centro-atlántica asciende continuamente magma fundido por la discontinuidad existente entre las placas divergentes, rellenándose así este vacío y creándose nuevo fondo oceánico. A medida que se enfría el magma saliente, aumenta en densidad y se hunde formando nueva litosfera que se agrega a la amplia protuberancia en expansión que constituye la dorsal. Allí donde dos placas colisionan, una puede sufrir subducción, es decir, hundirse bajo la otra, y ser reabsorbida en el manto. A lo largo de la línea que marca el descenso de la placa hacia el manto se



didades de 4000 metros. El agua no podía fluir hacia el norte, ya que las tierras existentes al norte de Gibraltar no habían empezado a separarse y la zona que comunicaba España con Africa carecía de la suficiente profundidad. Es probable que tampoco pudiera fluir el agua hacia el sur, porque los arrecifes coralinos que se habían formado sobre la plataforma de las Bahamas pudieron haber impedido la circulación entre el Atlántico y el Caribe. (El agua tampoco podía pasar desde el Caribe al Pacífico, porque zonas de lo que hoy es América Central se encontraban por debajo del nivel del mar. Hace 125 millones de años, América del Sur empezó a separarse de Africa y América del Norte continuó alejándose de Africa y España. La placa situada al norte de Venezuela sufrió subducción bajo la sudamericana a medida que el movimiento de América del Sur comprimía el Caribe.)

forma, en el fondo oceánico, una fosa profunda. La fusión de la placa en subducción genera una intensa actividad volcánica, a lo largo del borde de la placa suprayacente, que se localiza a unos 100 kilómetros de la fosa. Este tipo de actividad volcánica, junto con la colisión de los continentes, ha sido el responsable de la formación de la mayoría de las cordilleras de montañas. Al hundimiento de las placas que sufren subducción se debe, igualmente, la formación de la mayoría de los movimientos sísmicos que se producen en el mundo, convirtiéndose en puntos de referencia que marcan los bordes de las placas.

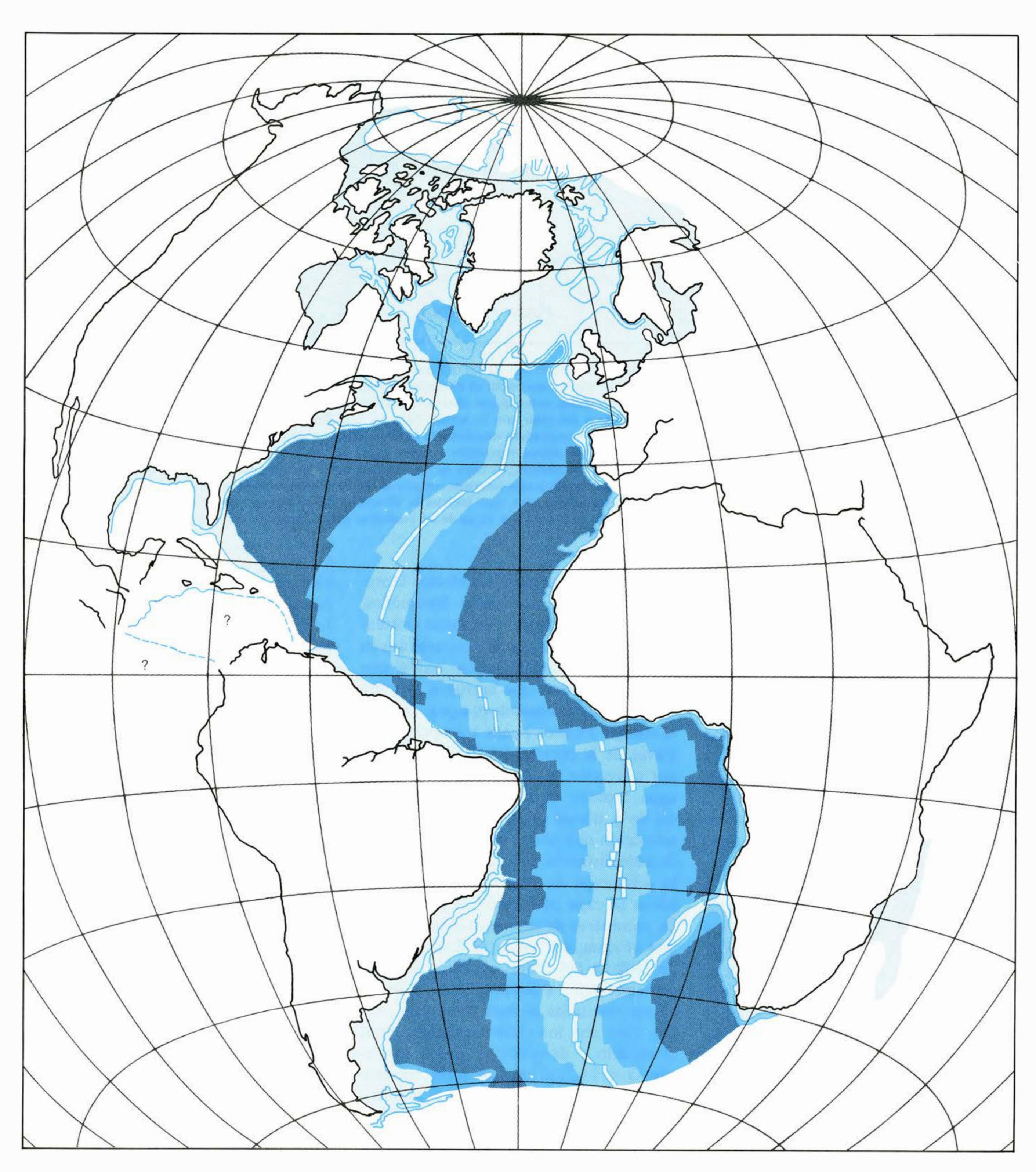
Se ha comparado el fondo del Atlántico con una gigantesca cinta transportadora que mueve la litosfera a una velocidad de centímetros por año, trasladándola desde los lugares donde se origina, en la dorsal centro-oceánica, hasta las fosas profundas donde desaparece. Esto hace que prestemos atención al hecho de



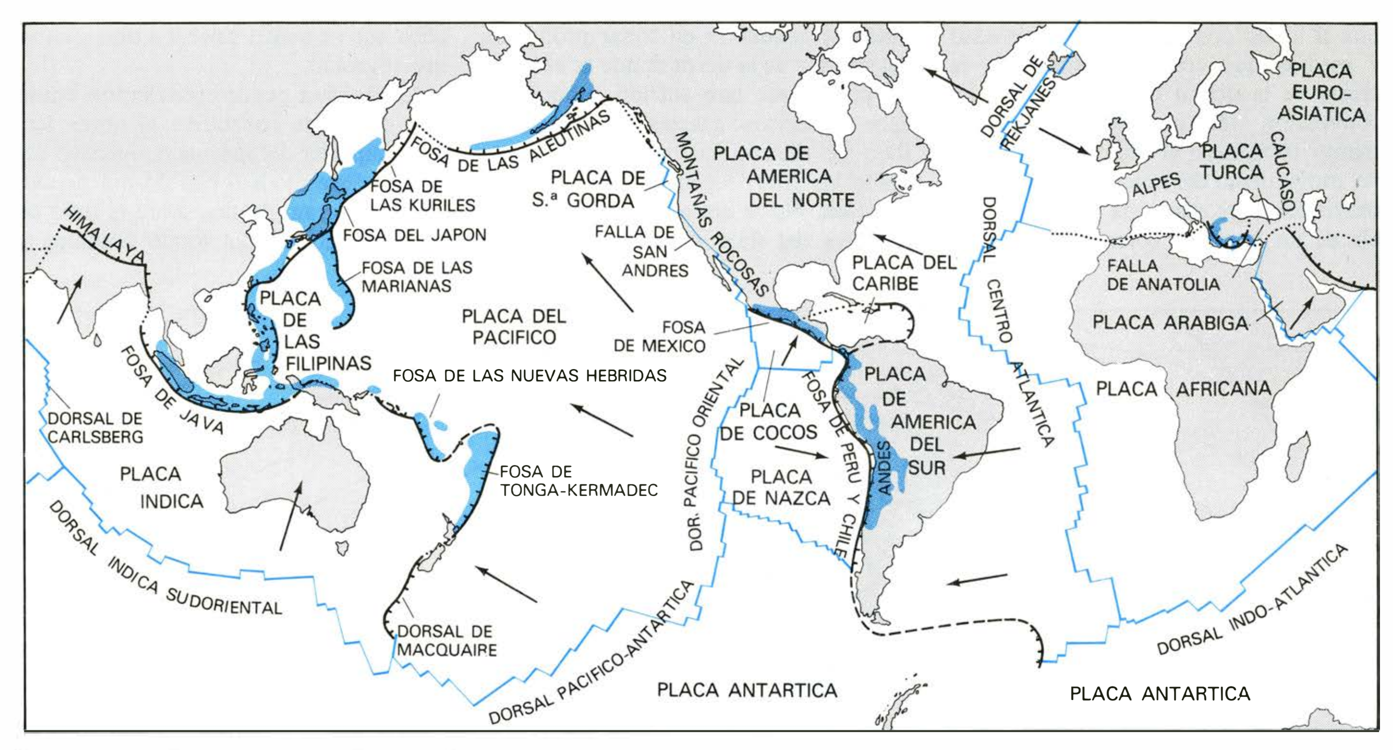
MAPA DEL ATLANTICO hace 80 millones de años (*izquierda*). Como puede apreciarse, el Atlántico Norte era ya todo un océano con profundidades inferiores a los 5000 metros. Su agua podía pasar a otros océanos a través del Caribe y del amplio Estrecho de Gibraltar. Groenlandia y América del Norte habían empezado a separarse, con lo cual se formaron, en esta etapa inicial, una serie de mares alargados y someros que se ramificaban a partir del cuerpo principal del Atlántico Norte. El Atlántico

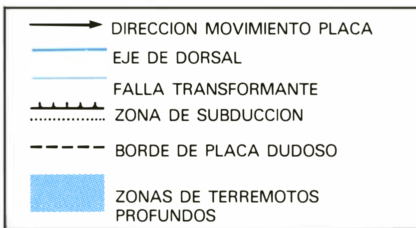
que el fondo oceánico aumente de edad a medida que crece la distancia a la cresta de la dorsal centro-oceánica. La convección térmica existente en el manto constituye el motor responsable del movimiento del fondo oceánico. Este movimiento proporciona la mitad visible de un bucle de convección; la otra mitad se encuentra en zonas profundas del interior de la tierra donde la masa de las placas que han sufrido subducción debe conservarse gracias a algún tipo de flujo de retorno de materia que vaya desde las fosas hacia la dorsal centro-oceánica. No se conocen todos los pormenores del flujo de retorno; ello explica que se hallen sujetos a una intensa investigación.

La litosfera puede considerarse como una capa que constituye el límite térmico superior del sistema convectivo del manto. Este modelo nos lleva a predicciones de gran alcance sobre el flujo de calor procedente del fondo oceánico y



Sur no se había desarrollado aún completamente. Con una profundidad máxima de sólo unos 4000 metros, estaba formado por la cuenca meridional y la septentrional, separadas entre sí por las dorsales de Río Grande y de Walvis. Las dorsales oponían rigurosa resistencia a la circulación entre ambas cuencas. Hace 36 millones de años (*derecha*), la mayoría de los rasgos topográficos más importantes del Atlántico se habían formado ya. El agua profunda podía circular en el Atlántico Sur gracias a que las dorsales de Río Grande y de Walvis habían sufrido una subsidencia suficiente. La diferencia principal que debemos reseñar entre entonces y ahora estriba en que España se encontraba bastante más alejada de Africa.





MOSAICO DE PLACAS RIGIDAS que constituye la litosfera terrestre, o envoltura más externa del planeta. Las placas, que están en continuo movimiento, se distinguen por su tamaño. Así tenemos desde las seis principales, como la que abarca virtualmente todo el Pacífico, hasta otras varias menores, como la que viene a coincidir con la actual Turquía. Las flechas indican los movimientos relativos de las placas cuando se considera estacionaria la placa africana. Las placas pueden interferirse de tres maneras distintas: por mutua separación a ambos lados de las dorsales centro-oceánicas; por colisión en los bordes de los continentes, en cuyo caso una placa sufre subducción, es decir, se hunde bajo la otra; y por presentar movimientos laterales recíprocos (en cuyo caso la fricción generada entre ellas da lugar a las fallas transformantes, que indican la dirección del movimiento entre dos placas.) La mayoría de terremotos se producen a lo largo de los contactos entre placas y a lo largo de las zonas de subducción.

sobre la profundidad del mismo. Las predicciones realizadas han resultado ser en gran parte verdaderas. Considérese la litosfera del Atlántico en un corte vertical. Originada a temperaturas elevadas en el centro de la dorsal centro-atlántica, la placa se enfría primero en su superficie superior, en donde está en contacto con el agua del mar que se encuentra a una temperatura de unos cero grados centígrados. A medida que el calor fluye a través de esta superficie y pasa al océano, la placa se enfría y aumenta de espesor. La descripción teórica de esta conducción térmica nos proporciona una ecuación simple que predice el flujo de calor que atraviesa la placa. De esa ecuación se obtiene la expresión, $q = 11{,}3/\sqrt{t}$, la cual nos indica que el flujo de calor (q) que pasa desde la placa al océano es una función de la edad (t) del fondo oceánico. Las dimensiones de q son las de la unidad de flujo de calor (10^{-6} calorías por centímetro cuadrado y por segundo) y las de t, un millón de años.

La validez de la expresión del flujo de calor puede ser comprobada midiendo q en distintos sitios del fondo oceánico y comparando los valores medidos con los valores predichos. Sin embargo, son difíciles de llevar a cabo las mediciones del flujo de calor debido a los efectos térmicos producidos por otros fenómenos. Las grietas verticales y las fallas existentes en el eje de la dorsal centro-atlántica fracturan el magma que se enfría con rapidez, de suerte que la nueva litosfera se hace muy permeable al agua. Por eso la energía térmica puede transportarse a través de la corteza del fondo oceánico por advección, es decir, por un flujo de agua movido térmicamente. Cerca del eje de la dorsal se pierde más calor por advección que por conducción. A considerable distancia del centro de la dorsal, la advección es despreciable. En esas zonas la capa permeable de litosfera se ha recubierto de sedimentos impermeables al agua; por tanto, el flujo de calor procedente del fondo oceánico se debe principalmente a conducción. Las mediciones realizadas en esas zonas concuerdan bien con los valores predichos por el modelo y demuestran que la expresión $q = 11{,}3/\sqrt{t}$ predice correctamente el flujo de calor que presenta la litosfera de edad inferior a 120 millones de años. En la litosfera más antigua, el calor adicional originado en el manto hace que el flujo de calor real sea mayor que el predicho por el modelo del límite térmico.

A medida que se enfría la litosfera, sufre una contracción térmica. La velocidad de la contracción vertical puede calcularse a partir del flujo de calor. Junto con el efecto de carga gravitatoria producido por el agua del mar, esta contracción es responsable de la profundidad del fondo oceánico. La teoría nos proporciona una expresión matemática muy sencilla: $D = 2500 + 350\sqrt{t}$, donde D simboliza la profundidad en metros y t, la edad del fondo en millones de años. Los datos empíricos indican que esta expresión, al igual que la correspondiente al flujo de calor, da valores correctos para las zonas litosféricas cuya antigüedad no sea superior a 60 millones de años. Tal como hemos dicho anteriormente, el eje de la dorsal centro-atlántica se encuentra 2500 metros por debajo del nivel del mar. La isobata (o curva de nivel que une los puntos de igual profundidad) de 3000 metros se halla sobre corteza oceánica cuya edad es de dos millones de años; la isobata de 4000 metros sobre la corteza cuya edad es de 20 millones de años y la isobata de 5000 metros sobre corteza cuya edad ya es de 50 millones de años.

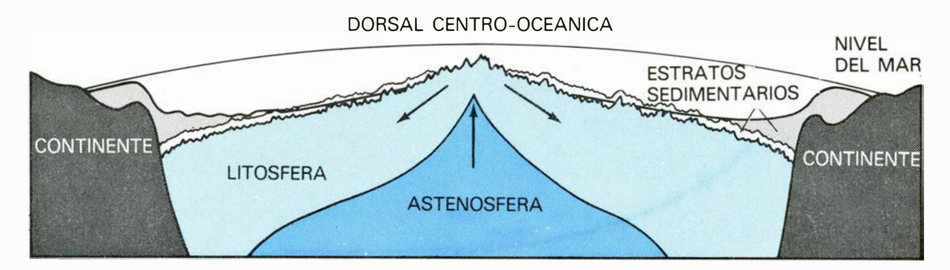
Las expresiones del flujo de calor y de la profundidad se desarrollaron a partir de un modelo que considera la litosfera como el límite térmico superior del sistema convectivo del manto. Un modelo

distinto introduce la hipótesis adicional de que el calor fluye desde el manto hacia la litosfera para mantener una isoterma (superficie de igual temperatura) de 1400 grados centígrados a una profundidad de unos 125 kilómetros. Bajo esa isoterma, el material del manto es menos sólido y, en consecuencia, el modelo es compatible con la teoría de la tectónica de placas que considera la litosfera como un material sólido (placas) que flota sobre un material que lo es menos. Este modelo de placas predice correctamente el flujo de calor y la profundidad de los fondos oceánicos, cualquiera que sea su edad. El modelo de la tectónica de placas postula un fuerte calentamiento procedente de las zonas sublitosféricas para explicar la menor velocidad de contracción del fondo oceánico de edad superior a los 60 millones de años. Una fuente de calor en el manto haría disminuir la velocidad de enfriamiento de la litosfera oceánica más antigua, motivando la disminución de la velocidad de contracción litosférica y responsable de que el fondo oceánico se encuentre a mayor profundidad.

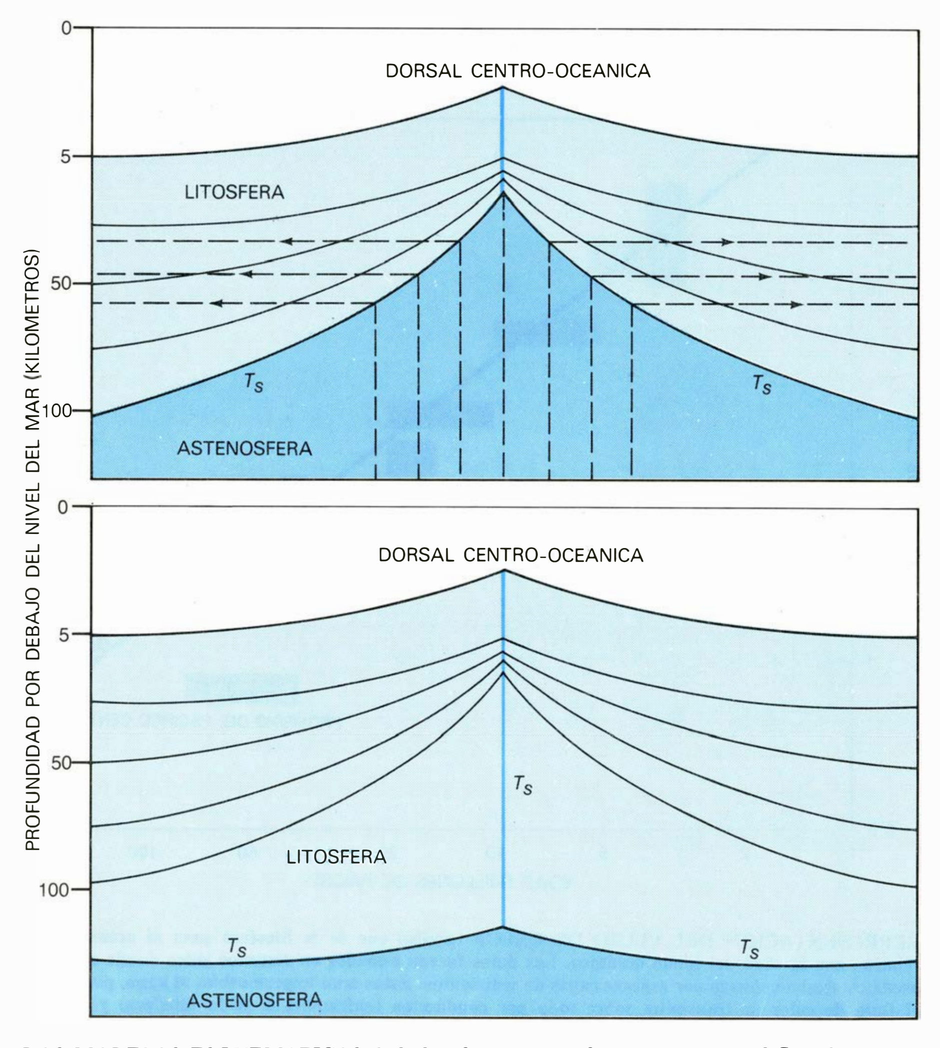
La velocidad de separación de las placas a ambos lados de la dorsal centro-oceánica puede determinarse a partir de las características magnéticas del fondo. A medida que el magma asciende hacia la dorsal, y se enfría formando fondo oceánico nuevo, se magnetiza según la dirección que el campo magnético terrestre tiene entonces. El campo geomagnético invierte su polaridad a intervalos irregulares de aproximadamente un millón de años; así, conforme el fondo oceánico va alejándose lentamente de la dorsal, se magnetiza dando una serie de largas bandas lineales de polaridad alternante que discurren paralelas al eje de la dorsal. Estas bandas de polaridad alterna producen anomalías magnéticas que perturban el campo local en un uno por ciento aproximadamente.

Los geofísicos han calculado la perturbación midiendo el campo alterado y restando de la medición el valor que el campo tendría en ausencia de las masas magnéticas locales. A partir de estos cálculos se ha determinado la anchura de las bandas polarizadas. Las inversiones del campo a lo largo de los últimos cinco millones de años han sido datadas a partir de las "firmas" que han quedado registradas en las coladas de lava emitidas en tierra firme. Estas dataciones señalan las edades de las bandas recientes del fondo oceánico; la velocidad de separación de dos placas puede calcularse directamente a partir de las edades y anchura de esas bandas. Ha sido posible, además, datar las inversiones del campo a lo largo de los últimos 150 millones de años, comparando las anomalías magnéticas de diferentes zonas del fondo oceánico. De este modo se han determinado las edades de todas las bandas del fondo oceánico.

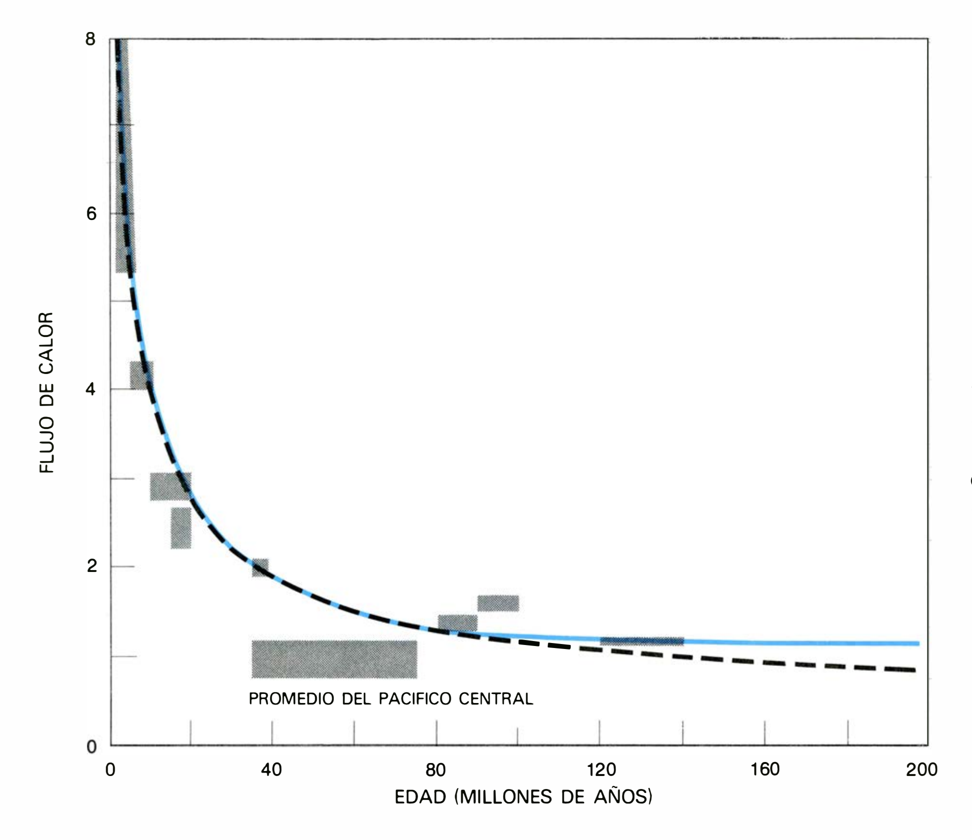
Los movimientos realizados en el pasado por el fondo oceánico pueden conocerse también gracias a las fallas transformantes producidas cuando dos placas contiguas se mueven una respecto de la otra. La fricción generada entre las placas origina intensas fuerzas de cizalladura que fracturan el fondo oceánico y dan lugar a los cañones. Las fallas transformantes son sísmicamente activas porque los esfuerzos producidos en las placas a lo largo de estos accidentes generan fuertes terremotos. Cuando se libera a las placas de esas tensiones, la cicatriz producida en la corteza oceánica

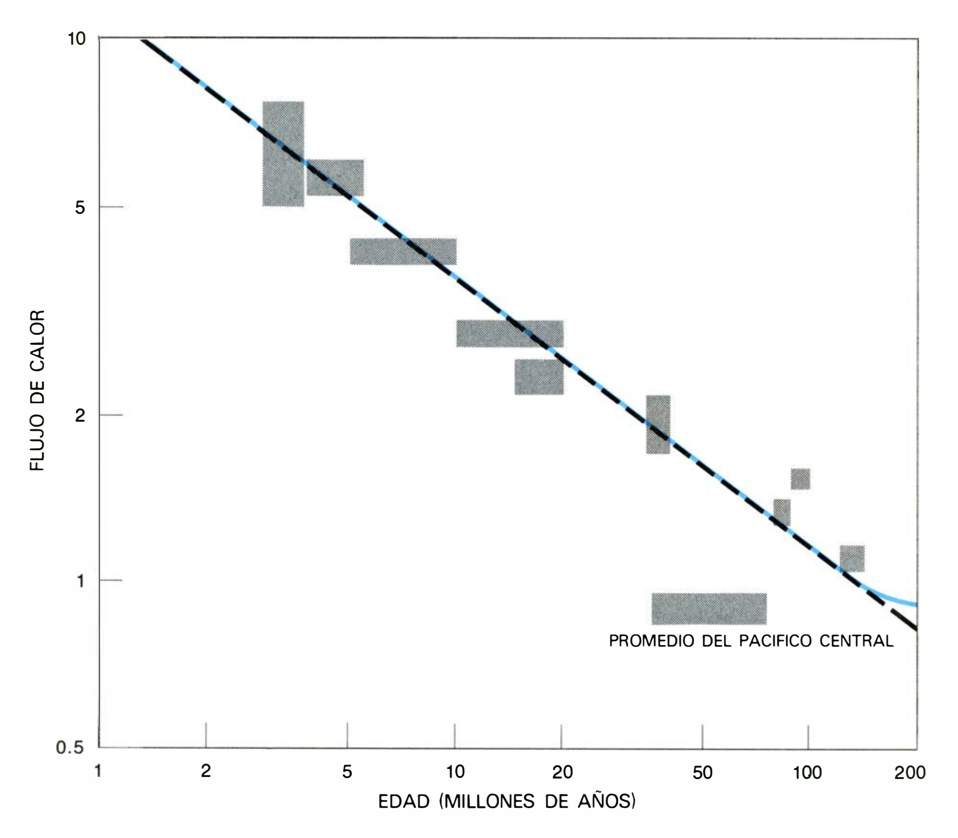


CORTE A TRAVES DE LA DORSAL CENTRO-ATLANTICA, que muestra cómo el fondo oceánico aumenta de profundidad, simétricamente, al crecer la distancia a la cresta. En las proximidades de los continentes los sedimentos aumentan de espesor. Los estratos sedimentarios están constituidos por arcilla, terrígenos (*gris claro*) y carbonato cálcico (*color claro*). La litosfera flota sobre la astenosfera.



DOS MODELOS ESQUEMATICOS de la litosfera y astenosfera que muestran el flujo de materia (*líneas discontinuas*) y las isotermas, o superficies de igual temperatura (*líneas de trazo continuo*). La temperatura del sólidus (T_s) es la isoterma que marca la transición desde el material sólido de la litosfera hasta el material parcialmente fundido que constituye la astenosfera. El modelo del límite térmico (*arriba*) considera la litosfera como el límite térmico superior del sistema convectivo del manto. El modelo de placa (*abajo*) introduce el supuesto adicional de que el calor fluye desde el manto hasta la litosfera para mantener una temperatura del sólidus de 1400 °C a una profundidad de unos 125 kilómetros. Las otras isotermas resultan casi idénticas en el modelo del límite térmico y en el de placa.





REPRESENTACION DEL FLUJO DE CALOR (*arriba*) **que de la litosfera pasa al océano y su relación con la edad del fondo oceánico. Los datos fueron tomados en distintos sitios donde el fondo oceánico estaba cubierto por espesas capas de sedimentos. Estos eran impermeables al agua, por lo que el flujo de calor se transmitía sobre todo por conducción (enfriamiento de la litosfera) y no por advección (transporte de energía térmica debido al movimiento del agua). Las zonas oscuras indican el intervalo de los datos medidos. La línea de trazos señala la relación existente entre flujo de calor y edad deducida a partir del modelo del límite térmico. El modelo predice que $q = 11{,}3/\sqrt{t}$, donde q simboliza el flujo de calor, en 10^{-6} calorías por centímetro cuadrado y por segundo, y t, la edad del fondo oceánico, en millones de años. Aparte del promedio anómalo correspondiente al Pacífico central, la ecuación puede aplicarse a toda la litosfera de edad inferior a los 120 millones de años. El modelo de placa que predice que $q = 0{,}9 + 1{,}6\ e^{(-t/62.8)}$ (*línea de color*) se cumple para cualquier edad. La representación según una recta (*abajo*) del logaritmo del flujo de calor en función del logaritmo de la edad del fondo oceánico muestra la relación $1/\sqrt{t}$ para los últimos 120 millones de años de historia**

permanece, pero ya no se comportará en adelante como zona sísmicamente activa. Este tipo de cicatrices son las zonas de fractura que rompen la simetría norte-sur del Atlántico. Así pues, mientras una falla transformante indica la dirección actual del movimiento entre dos placas, una zona de fractura indica la dirección del movimiento relativo que tuvo lugar en determinados momentos del pasado.

Estudiando las anomalías magnéticas y las zonas de fractura se pueden reconstruir las posiciones relativas de las distintas placas (y por tanto las de los continentes que las integran) en casi cualquier momento del pasado. Pero, en la práctica, los geofísicos deben limitarse a los últimos 200 millones de años. A medida que se forma una porción de fondo oceánico nuevo entre placas divergentes en la dorsal centro-oceánica a 2500 metros de profundidad se produce un envejecimiento, enfriamiento y contracción del mismo hundiéndose a zonas más profundas. Algunos millones de años después de que los continentes empezaran a separarse, el fondo oceánico de sus bordes se encontraba solamente a unos 3000 metros de profundidad. 50 millones de años más tarde, el fondo de estas zonas marginales se había hundido hasta una profundidad de 5000 metros como mínimo. La erosión de los continentes iba acumulando sedimentos terrígenos en las plataformas continentales y en los espacios adyacentes dejados por el fondo oceánico en continua depresión. El enorme peso de estos sedimentos hundió todavía más la corteza oceánica.

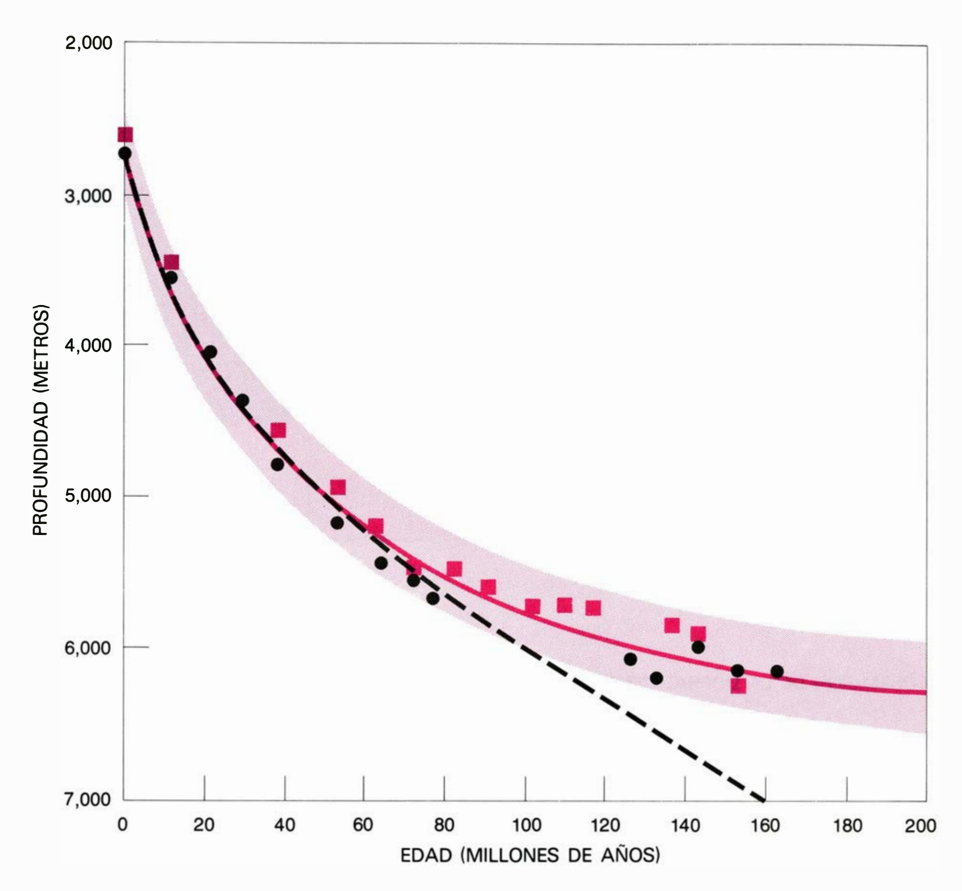
Antes de proseguir con la reconstrucción de la historia batimétrica del Atlántico, es necesario introducir un último concepto: la formación de las dorsales asísmicas (como la de Río Grande y la de Walvis) que constituyen zonas del océano anómalas por su poca profundidad. La velocidad de subsidencia de las dorsales asísmicas revela si actuaron o no alguna vez como barreras importantes respecto a la sedimentación y circulación de agua. Las dorsales asísmicas se van formando lentamente en zonas de magmatismo excepcionalmente activo, situadas sobre o cerca del eje de la dorsal centro-oceánica. El magma saliente forma apilamientos volcánicos que se levantan hasta el nivel del mar. A medida que la expansión del fondo oceánico va alejando uno de estos apilamientos de las zonas activas, éste se convierte en un mero pasajero pasivo y asísmico colocado sobre el fondo. Así pues, la dorsal asísmica, que no es otra cosa que una sucesión de apilamientos de este tipo, se hunde a la misma velocidad que

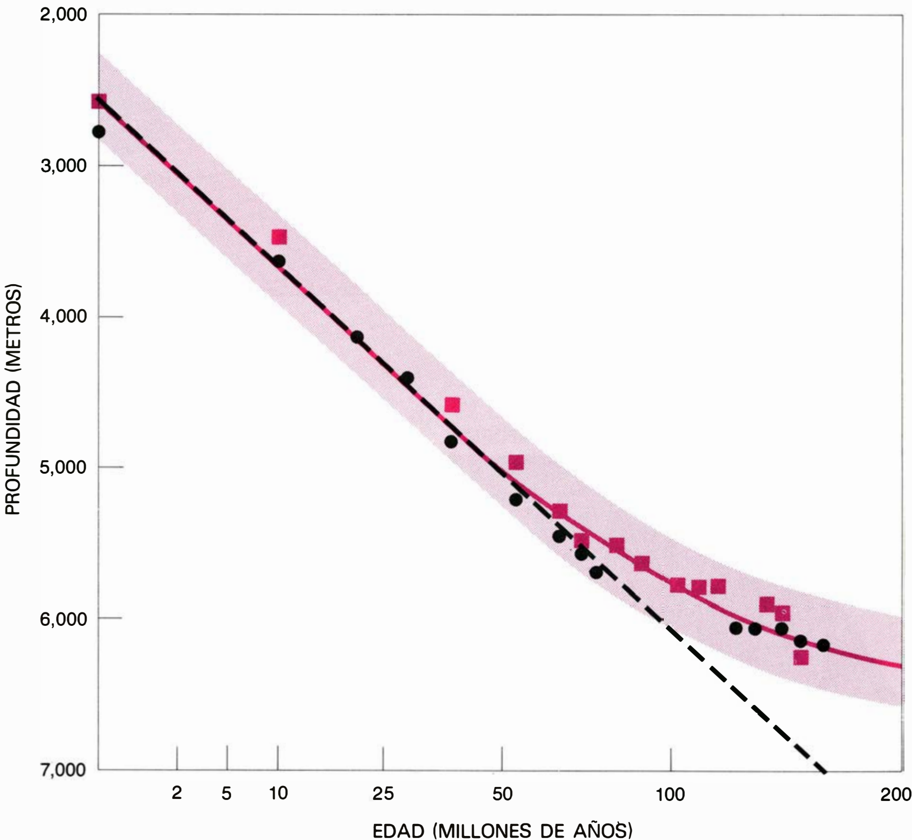
el fondo oceánico que la sostiene, manteniéndose siempre su cresta a unos 2500 metros por encima del fondo subsidente.

Los geofísicos han reconstruido la historia del Atlántico a partir de modelos térmicos, de la teoría de la tectónica de placas y de las mediciones realizadas en las zonas oceánicas profundas. A mediados del Jurásico, hace 165 millones de años, América del Norte, América del Sur, Groenlandia, Europa y Africa estaban unidas formando una gran extensión de tierra firme. En nuestra reconstrucción hemos realizado el encaje de los continentes utilizando como límite de los mismos las actuales plataformas continentales que están mucho más cerca de sus bordes antiguos que las actuales líneas costeras. Cabe destacar lo bien que los continentes han conservado su forma original. Si pudiéramos reunirlos ahora de nuevo, Groenlandia encajaría perfectamente con la línea de costa de Europa septentrional, y América del Norte rodearía la protuberancia africana y las tierras que constituyen lo que hoy en día son España y Francia. América del Sur, junto con la extensión correspondiente a la plataforma de Falkland, encajaría muy bien con el margen africano.

A mediados del Jurásico el magma fundido empezó a separar las tierras formando los continentes y originando una corteza oceánica entre América del Norte y Africa, separando los continentes septentrionales de Africa y América del Sur y creando los embriones del Atlántico Norte y del Caribe. Hace 125 millones de años el Atlántico Norte había desarrollado ya una dorsal centro-oceánica activa alcanzando profundidades elevadas, del orden de los 4000 metros. El agua de este nuevo océano apenas podía fluir hacia el norte, ya que las tierras situadas al norte de Gibraltar no habían empezado a separarse aún y el paso entre España y Africa era muy estrecho, quizás incluso se encontraba por encima del nivel del mar.

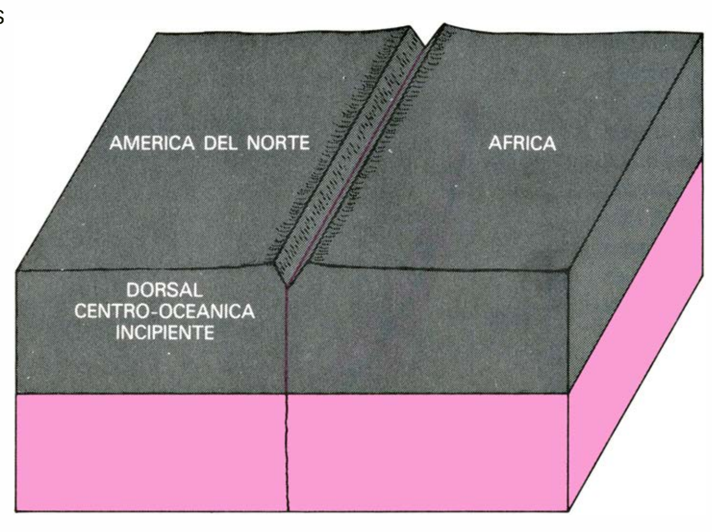
En esta época había partes de Centroamérica que se encontraban bajo el nivel del mar, pero hay pruebas de que el agua del Atlántico no podía pasar al Pacífico a través del Caribe. La antigua corriente marina entre el Atlántico y el Caribe dependía del desarrollo de la plataforma de las Bahamas que, saliendo de Florida, se dirige hacia el sudeste. Si los arrecifes coralinos crecieron sobre la plataforma a la misma velocidad que ésta se hundía, el Atlántico Norte pudo haber permanecido separado del Pacífico durante un considerable período de tiempo. De cualquier modo, la circulación entre el Atlántico y el Caribe debió ser casi tan limitada y lenta como la cir-



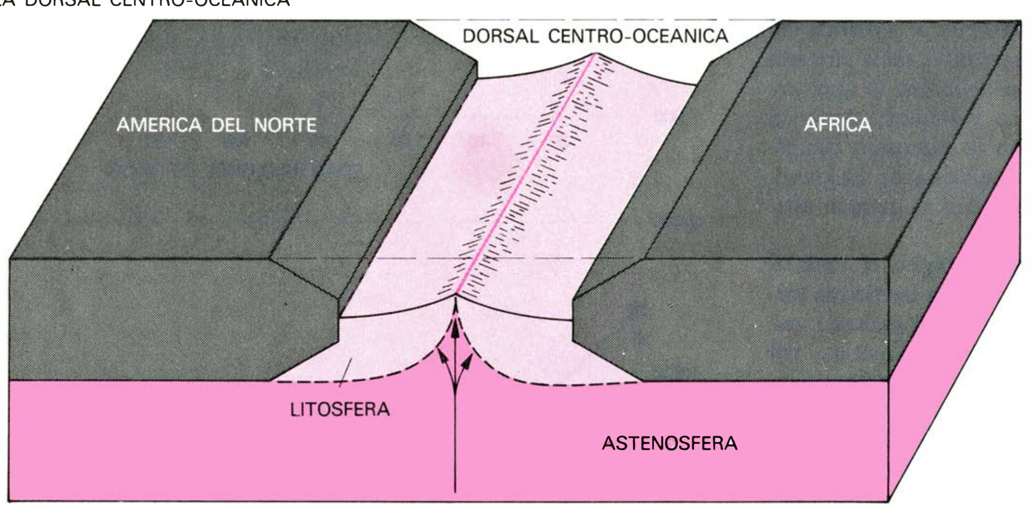


REPRESENTACION DE LA PROFUNDIDAD (*arriba*) del fondo oceánico en función de su edad realizada para el Atlántico Norte (*cuadrados de color*) y para el Pacífico septentrional (*círculos negros*). La zona coloreada representa la dispersión estimada para los datos de profundidad. El modelo del límite térmico predice que $D = 2500 - 350\sqrt{t}$ (*línea de trazos*), en donde D es la profundidad en metros y t la edad del fondo en millones de años. El modelo de placa predice que $D = 6400 - 3200\, e^{(-t/62.8)}$ (*línea de color*). La línea recta (*abajo*) que relaciona la profundidad con la raíz cuadrada de la edad pone de manifiesto la relación de $\sqrt{t}$ con el fondo oceánico de edad inferior a 60 millones de años. El modelo del límite térmico funciona con dificultad para todo fondo oceánico más antiguo. El modelo de placa proporciona valores correctos para la profundidad de un océano, igual que ocurría con el flujo de calor.

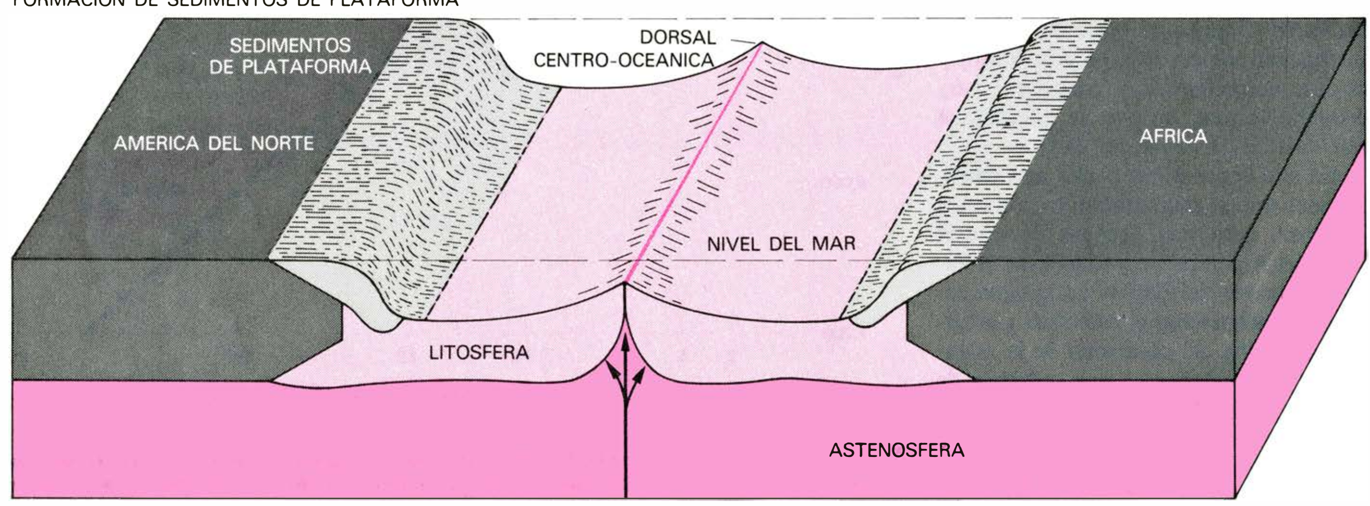
SEPARACION DE LOS CONTINENTES



FORMACION DE LA DORSAL CENTRO-OCEANICA



FORMACION DE SEDIMENTOS DE PLATAFORMA



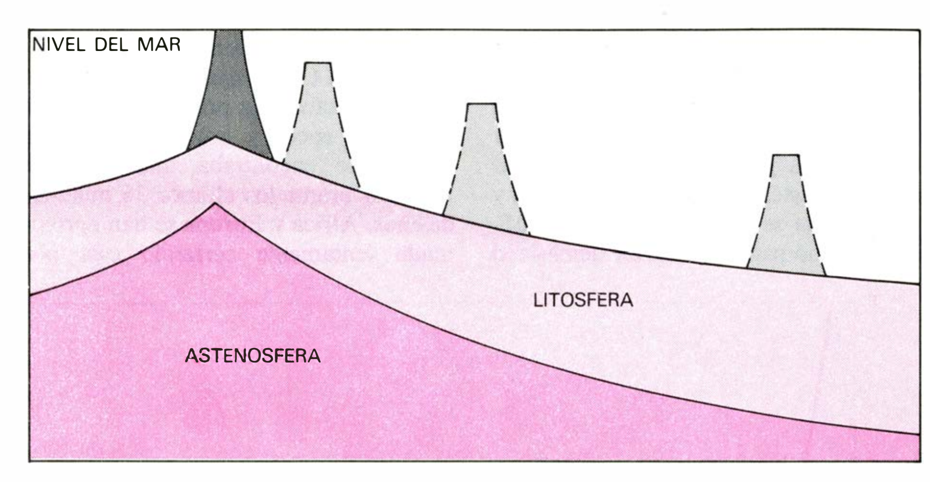
LA SEPARACION DE LOS CONTINENTES empieza en la incipiente dorsal centro-oceánica donde el magma fundido, procedente de las zonas profundas de la tierra, asciende y divide una extensión de tierra firme en continentes (*arriba*). En la dorsal, el magma fluye continuamente por entre la separación existente entre los continentes, haciéndoles alejarse (*centro*). A medida que el magma saliente va enfriándose, aumenta de densidad y se hunde formando litosfera nueva que se suma a la amplia protuberancia en expansión que constituye la dorsal centro-oceánica. Los productos resultantes de la erosión continental empieza a acumularse sobre las plataformas continentales (*abajo*). Estos diagramas son propiedad de Andrew Tomko.

culación existente en una masa de agua estancada. Este estancamiento pudo haber sido responsable de la deposición de niveles salinos a lo largo de las zonas marginales de Africa y América del Norte y de la preservación, en los sedimentos, de enormes cantidades de materia orgánica.

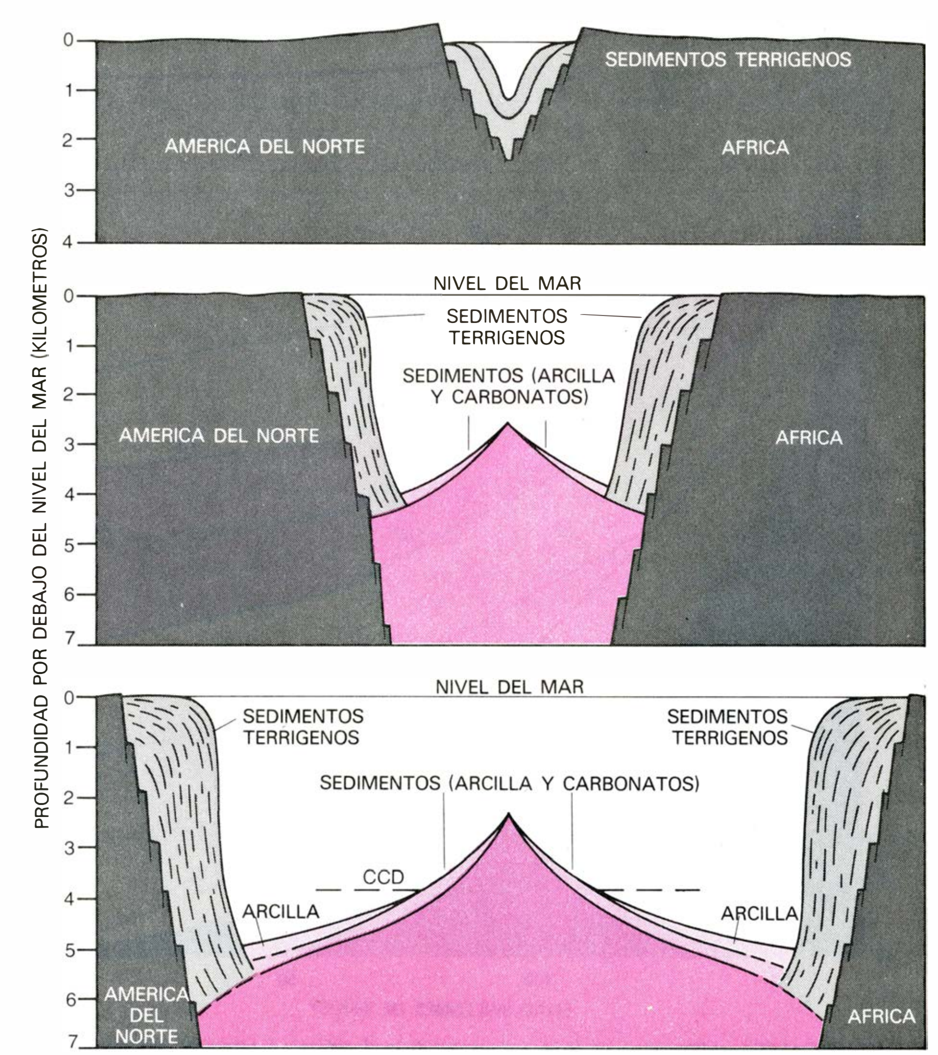
Hace unos 125 millones de años América del Sur empezó a separarse de Africa y España, junto con Africa, comenzó a alejarse de América del Norte. El desplazamiento de América del Sur respecto de América del Norte comprimió la región del Caribe provocando la subducción de la placa existente al norte de Venezuela. Hace 80 millones de años, el Atlántico Norte era ya un verdadero océano. Algunas de sus partes se encontraban a más de 5000 metros de profundidad y su agua podía circular hacia otros océanos a través del Caribe y del amplio Estrecho de Gibraltar. América del Norte y Groenlandia habían empezado a separarse formando en esta etapa inicial mares estrechos y de poca profundidad que eran apéndices de la masa principal del Atlántico Norte.

El Atlántico Sur es más reciente que el Atlántico Norte, habiendo empezado a desarrollarse tan sólo hace 80 millones de años. Dividido en las cuencas septentrional y meridional por las dorsales de Río Grande y de Walvis, el Atlántico Sur alcanzó una profundidad máxima de sólo unos 4000 metros. Las dorsales restringieron fuertemente la circulación entre las cuencas. La cuenca meridional estaba unida a otros océanos en su terminación sur, pero la cuenca septentrional se mantenía aislada debido a las protuberancias formadas por Africa y América del Sur. Estas protuberancias y las dorsales de Río Grande y de Walvis cercaron la cuenca septentrional durante los primeros 20 o 30 millones de años de su historia. Durante este período, el llenado intermitente y la subsiguiente evaporación condujeron a la deposición de sales en toda la zona. Estas sales se encuentran hoy sobre las plataformas continentales de Africa y de América del Sur al norte de estas dorsales.

Hace unos 65 millones de años-Groenlandia empezó a separarse de Europa. Hasta hace unos 20 millones de años, una dorsal asísmica cercana a Islandia había impedido el flujo de agua fría desde el Artico al Atlántico. En esta época, sin embargo, la dorsal había experimentado una subsidencia lo suficientemente importante como para permitir el paso de agua fría marina desde el norte hacia el Atlántico, produciendo la vigorosa circulación que el océano presenta en la actualidad.

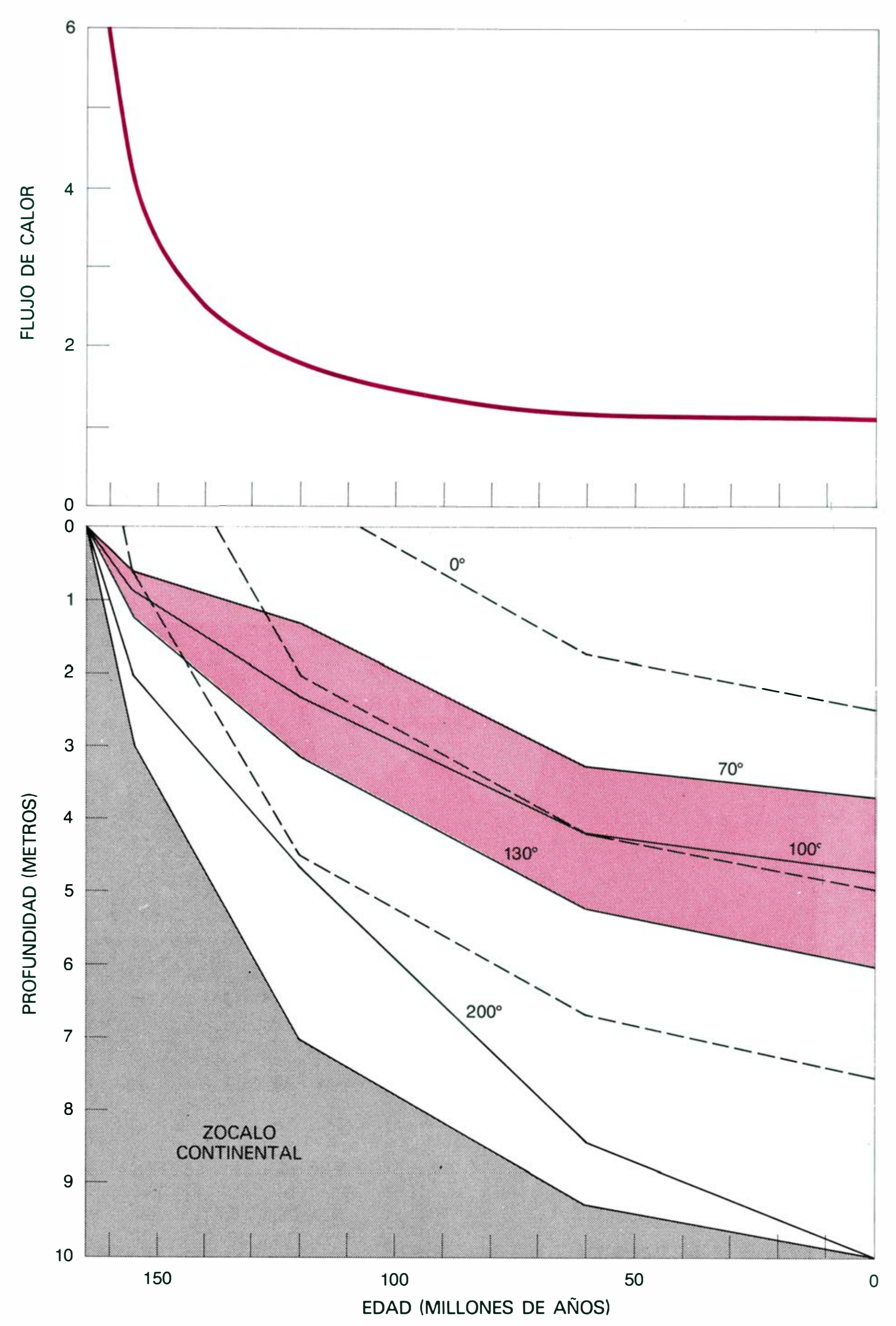


DORSALES ASISMICAS, creadas lentamente en zonas de magmatismo muy activo situadas sobre, o cerca, del eje de la dorsal centro-oceánica. El magma saliente forma un apilamiento volcánico (*región gris oscura*) que se levanta hasta el nivel del mar. A medida que la expansión del fondo oceánico mueve el apilamiento (*regiones en gris claro*), alejándolo de las zonas activas, éste se convierte en una especie de pasajero pasivo y asísmico colocado sobre el fondo. El apilamiento, o dorsal asísmica, se hunde a la misma velocidad que el fondo oceánico subyacente. Estas dorsales son barreras antisedimentarias.



MODELO DE SEDIMENTACION en tres etapas que ilustra el desarrollo del Atlántico Norte. Hace unos 165 millones de años (*arriba*) empezaron a acumularse sedimentos terrígenos, procedentes de los continentes, sobre las plataformas continentales en curso de formación. Hace 125 millones de años (*parte central*) el fondo oceánico se cubrió de arcilla roja (un material detrítico ligero procedente de los continentes) y de un barro calcáreo (carbonato cálcico derivado de la descomposición de partes esqueléticas de microorganismos). En las proximidades de la cresta de la dorsal centro-oceánica predomina el barro calcáreo. El denominado nivel de compensación del carbonato cálcico (CCD, del inglés "calcium carbonates compensation depth") se sitúa entre unos 4,5 y 5 kilómetros y es el nivel bajo el cual los sedimentos calcáreos se disuelven porque el agua del mar no está saturada de carbonato cálcico. Ello significa que, a medida que el fondo oceánico se hunde por debajo de la profundidad de compensación, dejan de depositarse sedimentos de carbonato cálcico (*abajo*). Sobre el fondo oceánico existente bajo el nivel de compensación sólo se sedimenta arcilla. La historia batimétrica informa sobre las plataformas.

La mayor parte de los rasgos topográficos del Atlántico se habían formado ya hace 36 millones de años. Tanto el Atlántico Norte como el Atlántico Sur eran anchos y profundos, aunque más aquél que éste. Las dorsales de Walvis y Río Grande se habían hundido lo suficiente como para no obstruir demasiado el flujo de agua profunda hacia el Atlántico Sur. El Caribe casi había adquirido su forma actual. La principal diferencia entre esta época y hoy en día es que España se encontraba más lejos de Africa. Durante los últimos 36 millones de años, Africa y Europa se han aproximado lentamente cerrando casi por completo la comunicación entre el Atlántico y el Mediterráneo.



FLUJO DE CALOR predicho a partir de los sedimentos de la plataforma continental en función de la edad (*arriba*). Se realizó suponiendo que el proceso térmico producido en las plataformas se parece al que tiene lugar en la dorsal. El flujo de calor disminuye rápidamente hasta el valor de equilibrio mantenido durante los últimos 50 millones de años. A partir de las predicciones de flujo de calor, y del conocimiento de la velocidad de subsidencia del fondo oceánico existente bajo los sedimentos, se puede reconstruir la historia completa de la evolución de la temperatura y la profundidad de los sedimentos existentes en cualquier punto de las plataformas continentales. El diagrama de la parte inferior muestra las isotermas (*líneas continuas*) y las características de la sedimentación (*líneas discretas*) en una zona hipotética situada sobre la plataforma continental externa de la parte oriental de Estados Unidos. Este tipo de reconstrucciones pueden tener serias repercusiones económicas. Uno de los principales factores considerados por los geólogos del petróleo, al determinar si los sedimentos de la plataforma continental contienen petróleo o gas, es su grado de madurez, es decir, el lapso de tiempo que los sedimentos han permanecido dentro de un intervalo óptimo de temperaturas (desde 70 hasta 130 grados C). Los sedimentos que se encuentran en este intervalo de temperaturas se representan aquí en color.

Inicialmente, esperábamos poder utilizar la anterior historia batimétrica del Atlántico para reconstruir cómo la morfología del fondo oceánico controló las corrientes de agua profundas y superficiales. Esperábamos poder relacionar directamente los cambios de la morfología del fondo con la distribución de los sedimentos. Nuestras esperanzas, demasiado optimistas, quizá no llegen nunca a realizarse, en razón de existencia de efectos locales de pequeña escala, como el flujo de agua a través de las zonas de fractura, que juegan un papel sumamente importante en la orientación de la circulación del agua oceánica. Con todo, la historia batimétrica puede servir de base para examinar el registro sedimentario en las zonas oceánicas profundas. A partir de la historia batimétrica podíamos determinar la distribución superficial de sedimentos de mar profundo a través del tiempo. A medida que el magma asciende entre los continentes y éstos se separan, se depositan sedimentos terrígenos en la plataforma continental, mientras que los sedimentos arcillosos y carbonatados se depositan en las zonas oceánicas.

Las dorsales asísmicas actúan a modo de barreras respecto a la llegada de sedimentos procedentes de las plataformas continentales. Cuando las dorsales presentan subsidencia aún controlan los tipos de sedimento depositados en la zona de mar profundo. Tal como hemos mencionado ya, el nivel de compensación del carbonato cálcico marca el nivel bajo el cual el agua marina está subsaturada de este compuesto. El nivel de compensación difiere en cada cuenca, variando incluso dentro de una misma en función de su situación. Si se conocen el nivel de compensación y su variación con respecto a su situación dentro de la cuenca, se puede reconstruir la distribución de los sedimentos de carbonato y arcilla a través del tiempo. Sobre una corteza oceánica que se encuentre por encima del nivel de compensación domina el carbonato cálcico, mientras que sobre la corteza existente por debajo del citado límite se formarán sedimentos constituidos sólo por arcillas de grano fino, barros silíceos y sedimentos hemipelágicos ligeros procedentes del continente.

La historia batimétrica del Atlántico explica el paradójico descubrimiento de sedimentos de carbonato cálcico por debajo del nivel de compensación, en la base de columnas de sondeos extraídos de los sedimentos próximos a las plataformas continentales por barcos equipa-

dos para realizar perforaciones en zonas de mar profundo. Cuando se forma la corteza oceánica, ésta se localiza por encima del nivel de compensación depositándose sobre ella los sedimentos de carbonato cálcico. Con el tiempo, el fondo se aleja del centro de expansión, hundiéndose finalmente hasta un nivel situado por debajo del nivel de compensación. Entretanto, los sedimentos de carbonato han sido recubiertos por barros arcillosos y silíceos que, si se encuentran lo bastante cerca de la costa, son recubiertos a su vez por los sedimentos terrígenos procedentes de la plataforma. Los terrígenos, la arcilla y los sedimentos silíceos protegen los depósitos de carbonato cálcico del agua del mar que los disolvería si estuvieran en contacto directo con ella a esas profundidades.

La historia batimétrica del Atlántico proporciona no sólo información sobre el fondo oceánico, sino también sobre las plataformas continentales. A partir de columnas de sedimentos obtenidas en las plataformas, los geofísicos han concluido que estas últimas presentan una subsidencia que se produce aproximadamente a la misma velocidad que la correspondiente a la dorsal centro-oceánica. El parecido entre las velocidades de subsidencia se ha atribuido a un proceso térmico que se produce en la plataforma y que recuerda el proceso térmico presente en la dorsal. Si esto es así, la plataforma debería perder calor al mismo ritmo que el fondo oceánico circundante. Puesto que la mayor parte de los sedimentos depositados en las plataformas son impermeables al agua, el calor allí existente se disipará sólo por conducción y no por advección. En consecuencia, la ecuación que describe el flujo de calor en función de la edad de la dorsal centro-oceánica sirve para las plataformas continentales introduciendo sólo ligeras modificaciones. La temperatura del sedimento, a una profundidad cualquiera, es directamente proporcional al flujo de calor multiplicado por la profundidad y dividido por la conductividad. Así pues, si se conoce el espesor de los sedimentos y su velocidad de sedimentación, se puede reconstruir toda la historia de la evolución de la temperatura y profundidad de cualquier lugar de la plataforma.

Estas historias pueden tener consecuencias económicas importantes. Uno de los factores principales que determinan si los sedimentos contienen o no petróleo o gas es su grado de madurez, es decir, el lapso de tiempo durante el cual los sedimentos han estado sometidos a temperaturas situadas dentro de un determinado intervalo óptimo.

Las microcinas

Constituyen un grupo de antibióticos producidos por bacterias presentes en el intestino del hombre y los animales; es probable que estén implicados en la colonización microbiana normal y patológica del tracto digestivo

Carlos Asensio y Fernando Baquero

Los seres vivos más abundantes de la biosfera en número y en biomasa absoluta son los microorganismos, que, sin embargo, resultan invisibles para el ojo humano. Debido a este hecho, no es fácil percibir la omnipresencia, la trascendencia y las relaciones de estos organismos inferiores (virus y bacterias, algas, protozoos, hongos y otros organismos microscópicos) con los demás seres vivientes, incluidos nosotros mismos. Por razones obvias, el hombre se ha preocupado de manera particular de aquellos microorganismos causantes de enfermedades, que constituyen una mínima fracción del mundo microbiano. Así, la cultura popular se refiere con cierta familiaridad al virus de la gripe o de la viruela, al gonococo o al bacilo de Koch, por ejemplo. Pero desconoce, en general, la gran contribución de los microorganismos en el ciclo de la materia, en las transformaciones geoquímicas o en la purificación de las aguas, en la formación de los suelos fértiles, o como estadios trascendentales de la cadena alimentaria en las tierras y los mares. Igualmente se desconoce la ecología microbiana de nuestro propio intestino, con más de 600 especies microbianas distintas, de enorme importancia para la comprensión de gran número de procesos fisiológicos y patológicos (desde la producción de vitaminas a la malnutrición, desde la protección frente a diarreas epidémicas hasta el cáncer de colon o el coma hepático). Y se desconoce no sólo desde la cultura popular, sino también –y esto es más grave–, en gran parte, desde la perspectiva de los científicos.

La microbiología es una ciencia que nació en el siglo XVIII con el holandés Antonie van Leevenhoek, inventor del primer microscopio, pero recibió su mayor impulso en la Europa de la segunda mitad del pasado siglo gracias al talento y al esfuerzo de hombres como Louis Pasteur, Robert Koch o Martinus Beijerink. Sin embargo, la ciencia microbiológica comienza ahora una nueva etapa al enfrentarse con la noción y el conocimiento de la comunidad microbiana natural, superando así una tradición ya secular en que se ocupó esencialmente de la clasificación taxonómica de las distintas especies microbianas y en la investigación fisiológica y patológica de estos organismos aislados en cultivos puros, es decir, estudiando cada uno de ellos aislados de otras especies, y utilizando además medios artificiales de crecimiento.

Este nuevo enfoque de carácter ecológico nos está acercando a la verdadera complejidad de los hábitats microbianos naturales, en los que entran en juego factores de interacción comunitaria que justamente ahora empezamos a explorar.

El ser humano constituye un buen ejemplo de soporte y ente asociado a una serie de ecosistemas microbianos complejos y diferenciados con los que convivimos normalmente, salvo alteraciones ocasionales, los cuales se localizan en distintas regiones de nuestras mucosas y epidermis. De estas comunidades, la más abundante y notoria es sin duda la que habita en el intestino grueso, en donde convive un mosaico variadísimo de especies y razas bacterianas unicelulares que constituyen en conjunto una población cuyo número supera ampliamente el total de las células de nuestro propio cuerpo.

La colonización microbiana de este hábitat comienza inmediatamente después de nuestro nacimiento. El primer contenido intestinal, el meconio, es estéril, pero su espacio es pronto invadido por una sucesión de distintos microorganismos que se multiplican activamente desplazando a los invasores anteriores. Al cabo de varios días, el medio intestinal alberga ya una población heterogénea de bacterias de distintos tipos que van integrando un ecosistema de complejidad creciente, hasta que llega a alcanzar una "comunidad climax", integración final que puede ser asimilada al concepto "sistema microbiano intestinal normal", puesto que cumple los postulados de Brock para la definición de todo sistema integrado: *homeostasis* (tiende a mantenerse íntegro frente a perturbaciones exógenas), *evolución* (la integridad se preserva en el tiempo), *mecanismos de defensa y reparación* (si una parte se pierde, tiende a reconstruirse) y de *reproducción* (un hábitat idéntico se colonizará por un sistema microbiano idéntico).

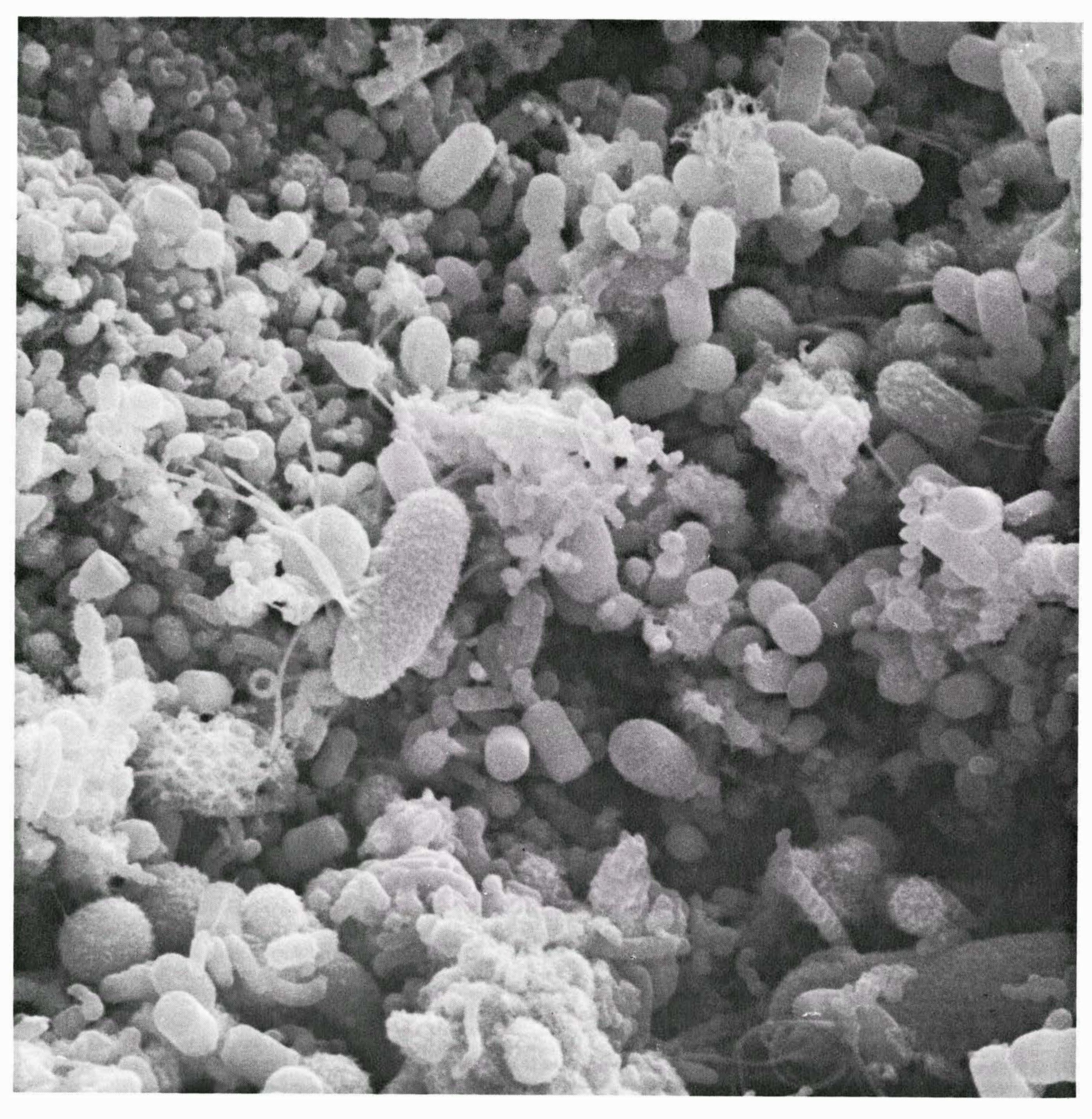
Es necesario concebir este sistema integrado en un estado dinámico. Si bien algunas poblaciones de la comunidad integrada parecen estar establecidas de forma casi permanente, otras se mantienen en un estado de flujo o sucesión, al poderse ocupar su nicho por microorganismos muy afines entre sí y de idéntico valor para la comunidad (diferentes variedades de *Escherichia coli*, por ejemplo). La ruptura de la integración del sistema es causa de patología, como ocurre con las infecciones intestinales agudas, en los síndromes de malabsorción o en el curso del tratamiento con antibióticos de amplio espectro.

En realidad, el sistema microbiano intestinal es un conjunto de elementos en interacción. El estudio de la interrelación entre microorganismos ha de ser, pues, la vía para la descripción y comprensión del sistema y sus propiedades. Desconocemos básicamente cuestiones tan importantes como los mecanismos respon-

NOTA DEL EDITOR:

Los autores desean agradecer la contribución al trabajo descrito en este artículo aportada por: A. Aguilar, A. Candela, M. R. Creixems, A. F. Duro, A. F. Jorge, C. Fernández, C. Granja, M. C. Martínez, A. Montes y J. C. Pérez-Díaz.

sables de la homeostasis, que impiden por ejemplo la multiplicación y asentamiento de los microorganismos extraños, inocuos o patológicos, que llegan constantemente al intestino desde el medio exterior. Desconocemos también de qué forma el sistema regula el tamaño de sus propias comunidades microbianas, cuyo desbordamiento es causa de patología. Es posible que los mismos mecanismos cumplan ambas funciones, y aún permitan de algún modo las fluctuaciones y sucesiones poblacionales típicas del ecosistema. Parece evidente que ciertos factores de éste contribuyen en alguna medida al control del crecimiento de unas y otras especies. Entre ellos están los condicionamientos nutricionales, la presión parcial de oxígeno, *p*H, factores biológicos y mecanismos procedente del huésped, etc., pero ninguno de ellos, de forma aislada o concertada, puede explicarnos por ejemplo la gran especificidad biológica de las sucesiones bacterianas ni la rapidez con que a menudo ocurren éstas, de forma que alguno de sus habitantes principales, tal como el colibacilo, son capaces de eliminar (o ser eliminados) con gran eficacia a otros organismos, incluso pertenecientes a la misma especie. Todo parece indicar que estas características podrían explicarse por un fenómeno de amensalismo, es decir, por la operación de agentes antibióticos segregados al medio por las especies indígenas, que interfieren selectivamente con el desarrollo de otros microorganismos, si bien eventualmente alguno de éstos podría establecerse con éxito en el sistema utilizando unos recursos similares, dando lugar a sucesiones o cambios



LA COMPLEJIDAD DE LOS ECOSISTEMAS MICROBIANOS del tracto digestivo de los mamíferos se refleja en esta microfotografía obtenida de una preparación del rumen bovino, por la técnica de microscopía electrónica de barrido. Esta fotografía, que muestra la gran cantidad y variedad de los microorganismos que conviven sobre la pared interior de aquel órgano, fue obtenida por R. P. Mc Cowan, K. J. Cheng y J. W. Costerton.

habitualmente sin trascendencia, pero que ocasionalmente condicionaría la aparición de patología. El mismo mecanismo serviría para controlar también el tamaño de las poblaciones productoras de esos agentes antibióticos, a los que serían sensibles por encima de ciertas concentraciones.

Se ha pretendido que las bacteriocinas podrían protagonizar estas competencias amensalistas. Las bacteriocinas son agentes letales de naturaleza proteica y tienen, por ello, un elevado peso molecular. Son producidas por bacterias de distintos tipos y las mejor conocidas son las denominadas *colicinas*, sintetizadas por un habitante característico de nuestra flora intestinal, la bacteria *Escherichia coli*, que recibe también el nombre común de colibacilo. Las colicinas conocidas –alrededor de una docena– son extraordinariamente específicas en su acción antibiótica, de forma que actúan únicamente sobre otras especies muy relacionadas con ésta, pero no afectan a la propia cepa productora. Las pruebas actualmente disponibles parecen descartar a las bacteriocinas como agentes homeostáticos principales del ecosistema intestinal. Así, cepas productoras de colicinas y otras sensibles a éstas coexisten perfectamente bien en el intestino, e incluso pueden predominar las últimas. De hecho, las colicinas apenas se producen en un medio tan pobre como el del intestino grueso. Por añadidura, debido a su naturaleza proteica las bacteriocinas en general son pronto inactivadas por los enzimas hidrolíticos denominados proteasas, cuya actividad es notoria en el contenido intestinal.

SIEMBRA DE *E. COLI*

INOCULACION DE EXTRACTO DE HECES

INCUBACION A 37 °C

DETECCION DE LAS MICROCINAS, obtenida a través de un procedimiento simple: sobre placas de cultivo sólido de agar, conteniendo un medio de sales minerales y glucosa, se sembró en toda su superficie una cepa de la bacteria *E. coli*, que serviría como indicador de antibiosis. Sobre la superficie del agar se depositó seguidamente una película de celofán y, sobre ésta, se inocularon en varios puntos distintos microorganismos entéricos aislados de heces infantiles. La fotografía de la izquierda muestra una de estas placas tras incubar a 37 °C por varias horas. Es evidente que, en este ejemplo, tres de las seis inoculaciones han excretado durante la incubación algún agente antibiótico que impidió el crecimiento a su alrededor de la cepa de *E. coli* indicadora, lo cual se visualiza por la aparición de un halo circular y transparente, en cada caso. La fotografía de la derecha corresponde a la misma placa después de haber separado la película de celofán (*abajo*), sobre la que pueden verse las colonias correspondientes a los seis inóculos entéricos. El uso del celofán permitió excluir a las bacteriocinas como causantes de las antibiosis detectadas, ya que por su gran tamaño molecular no pueden pasar a su través.

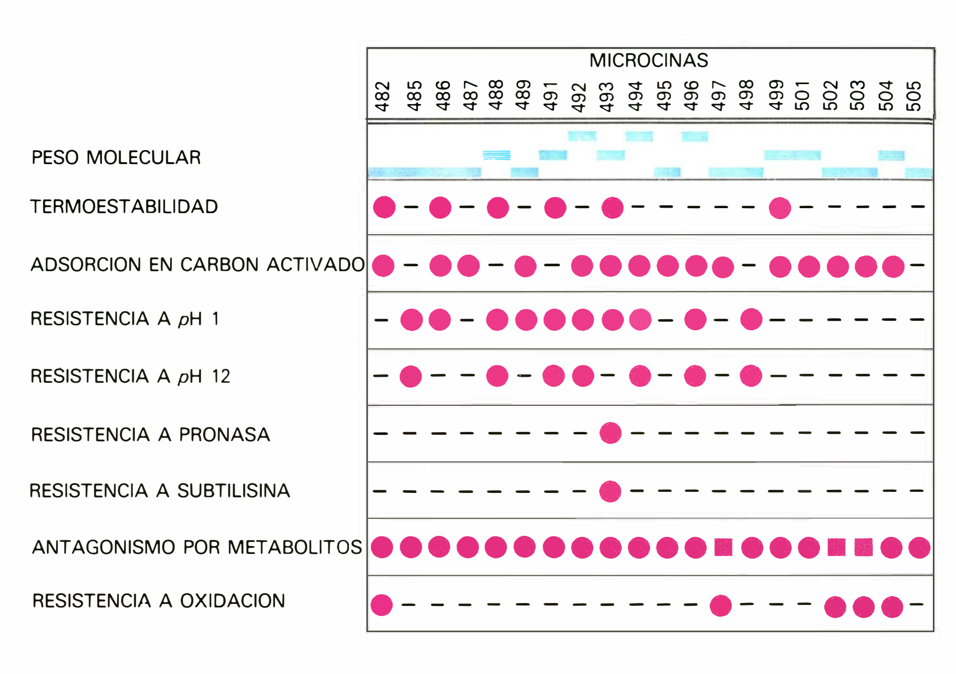
Por éstas y otras razones, llegamos al convencimiento de que en el ecosistema bacteriano del intestino grueso podrían existir agentes antagonistas de origen bacteriano distintos de las bacteriocinas y dotados de unas propiedades que les permitieran homeostatizar el sistema en la forma en que realmente ocurre. Para ello, imaginamos, deberían producirse en medios pobres y poseer estructuras de bajo peso molecular resistentes a las proteasas.

Con base a este planteamiento diseñamos un método de detección sencillo para explorar la posible existencia de substancias antibióticas con dichas propiedades. Con la aplicación de este método en una serie de microorganismos aislados de muestras fecales, pronto obtuvimos resultados en apoyo de la hipótesis de trabajo. Para nuestra sorpresa, más del 10 por ciento de las células ensayadas mostraron ser capaces de producir y excretar potentes agentes antibióticos de bajo peso molecular, a los que hemos denominado genéricamente con el nombre de *microcinas*.

El método de ensayo consistió en inocular sistemáticamente microorganismos entéricos aislados de heces infantiles sobre placas conteniendo un medio de cultivo mínimo, con sólo glucosa como nutriente orgánico. Estas inoculaciones, normalmente seis por placa [*véase la ilustración de la izquierda, arriba*], se aplicaron puntualmente sobre una película de celofán adherida sobre la superficie del propio medio de cultivo. Este había sido previamente sembrado en toda su extensión con una cepa de *E. coli* añadida para servir de indicador de cualquier antibiótico que pudiera excretar cualquiera de los seis inóculos puntuales bajo ensayo. Si tras incubar las placas a 37 °C durante 24 horas no se hubiese producido ninguna excreción antibiótica, la cepa indicadora crecería sobre toda la superficie del cultivo semisólido de la placa, mostrándose a simple vista con una opacidad blanquecina característica. Por el contrario, si se segrega algún antibiótico desde cualquiera de las inoculaciones puntuales, éste impedirá el crecimiento de la cepa indicadora alrededor de la colonia productora del antibiótico, lo cual se visualiza como un halo circular transparente sobre el fondo blanquecino que indica crecimiento en el resto de la placa. En la fotografía que ilustra este método, tres de las seis cepas entéricas ensayadas muestran claramente que durante su crecimiento se ha producido la segregación de algún agente difusible a través

del celofán que impidió el desarrollo de la cepa indicadora en un cierto radio de acción.

Con el método de ensayo descrito, puede anticiparse que las actividades antibióticas detectadas no son debidas a bacteriocinas convencionales, ya que, en primer lugar, éstas no se sintetizan virtualmente en medios de cultivo mínimo, y, además, porque dado su elevado peso molecular (por encima de 40.000) no podrían atravesar la película de celofán sobre la que se inoculan las cepas para ensayo, ya que sus poros impiden el paso de moléculas cuyo peso esté por encima de 15.000. Con este simple recurso, se logra obviar por un lado la posible interferencia por bacteriocinas, mientras que por otro se facilita la separación e identificación de antibióticos cuyo peso molecular podemos anticipar que será relativamente bajo, al menos inferior a 15.000.



IDENTIFICACION Y CARACTERIZACION de las microcinas. Se lleva a cabo realizando una serie de pruebas sistemáticas, tal como se ilustra en esta figura, que incluye la respuesta obtenida con 20 microcinas producidas por distintas células. Las barras azules situadas en la primera línea indican que el peso molecular de la correspondiente microcina es menor de 500 (*barras en la parte inferior*), entre 500-1000 (*medio*) o mayor de 1000 (*superior*). Las rayas a modo de guión indican que las microcinas no son significativamente inactivadas por los distintos tratamientos; los círculos rojos, que sí lo son. El cuadrado de color indica que el correspondiente antibiótico es antagonizado por *l*-metionina.

La mayor parte de las actividades antibióticas detectadas mostraron ser recuperables, en mayor o menor grado, en los sobrenadantes de medios de cultivo líquido en donde crecieron las distintas cepas productoras. Después de concentrar esos caldos a vacío y de precipitar las sales y otros componentes macromoleculares con metanol, se obtuvieron unos preparados con una mayor potencia antibiótica relativa, que sirvieron de base para una serie de pruebas bioquímicas y microbiológicas que nos permitieron diferenciar, identificar y caracterizar en primera aproximación las microcinas detectadas por la técnica del celofán.

Los datos obtenidos hasta el presente sobre medio centenar de microcinas, nos permiten perfilar unas propiedades generales que tipifican a estos compuestos como (salvo excepciones) de bajo peso molecular, en su mayoría por debajo de 500, y solubles en mezclas de metanol y agua; la mayoría son también termorresistentes y no hidrolizables por proteasas. En bastantes casos (20 por ciento) el efecto antibiótico es antagonizable por el aminoácido natural *l*-metionina, constituyendo éstos los únicos ejemplos encontrados en que las microcinas funcionan como antimetabolitos. Volveremos sobre este tipo de microcinas más adelante.

En ocasiones, los perfiles bioquímicos son aparentemente los mismos para más de una microcina, lo que sugiere la misma identidad, aunque se hubiesen aislado de cepas distintas. Nosotros aceptamos al principio tal identidad en una clasificación preliminar de estos agentes, pero hubimos de modificar este criterio varias veces al cotejar los resultados obtenidos en el espectro de especificidad antibiótica de las mismas preparaciones, en donde se obtienen resultados disimilares para microcinas aparentemente idénticas.

El espectro de acción antibiótico de las actividades detectadas hasta el presente, ofrece resultados de un gran interés. Este espectro se obtiene aplicando preparaciones parcialmente purificadas de las distinta microcinas sobre la superficie de cultivos sólidos en las que se ha sembrado un cierto tipo de microorganismos, para tratar de ver si su crecimiento es inhibido o no, y en qué medida en caso afirmativo, por cada una de las microcinas ensayadas. En todo caso empleamos 60 tipos distintos de microorganismos como indicadores de antibiosis, los cuales incluyen una serie de especies representativas de la familia de las Enterobacteriaceae (a la que pertenecen, como luego veremos, las cepas productoras de microcinas), más otras especies pertenecientes a tipos de bacterias gram-negativas no-fermentantes, bacilos y cocos gram-positivos y, finalmente, células eucarióticas (no-bacterianas) unicelulares, del grupo de las levaduras. Los resultados obtenidos indican que las microcinas son particularmente activas sobre las especies más próximas a las productoras de esos agentes, lo que refleja la competencia real en el mundo microbiano entre especies afines que tienden a ocupar el mismo nicho. Sin embargo, se mostraron también activas sobre otros tipos de organismos, incluso tan lejanos filogenéticamente como son las levaduras. Este último hallazgo constituyó una sorpresa realmente inesperada, ya que nos muestra una propiedad no compartida por los antibióticos usuales.

Desde el principio de nuestro trabajo nos interesamos por conocer la sensibilidad de las cepas productoras de las microcinas a sus propios productos. La respuesta, con carácter general, es que son insensibles cuando se aplica la técnica del celofán, pero son inhibidas en su crecimiento cuando se emplean preparaciones más potentes del correspondiente antibiótico. Esta observación podría significar la existencia de un mecanismo de autorregulación ecológica, como se discutirá más adelante. Asímismo, hemos estudiado en algún caso la potencia antibiótica de las microcinas en comparación con otros agentes convencionales utilizando el índice denominado Concentración Mínima Inhibitoria (CMI) contra una selección de microorganismos causantes de infecciones hospitalarias. El valor de este índice para la microcina 15, por ejemplo, llega a ser tan bajo como 0,01 microgramo/mililitro para *E. coli y Citrobacter*, lo que sitúa a ciertas microcinas entre los antibióticos de mayor actividad sobre estas especies.

¿Cuántas microcinas distintas podría

CEPAS ENSAYADAS		MICROCINAS 482	485	486	487	488	489	491	492	493	494	495	496	497	498	499	501	502	503	504	505	506	507	508	509	510
ENTEROBACTERIACEAE																										
E. COLI	(6)																									
CITROBACTER	(2)																									
KLEBSIELLA	(4)																									
ENTEROBACTER	(2)																									
SERRATIA	(1)																									
ERWINIA	(1)																									
PROTEUS	(5)																									
SHIGELLA	(3)																									
SALMONELLA	(7)																									
GRAM (–) NO FERMENTANTES																										
ACINETOBACTER	(1)																									
PSEUDOMONAS	(7)																									
GRAM (+) BACTERIAS																										
BACILLUS	(1)																									
STREPTOCOCCUS	(5)																									
STAPHYLOCOCCUS	(6)																									
MICROCOCCUS	(1)																									
LEVADURAS																										
CANDIDA	(5)																									
RHODOTORULA	(2)																									
SACCHAROMYCES	(1)																									

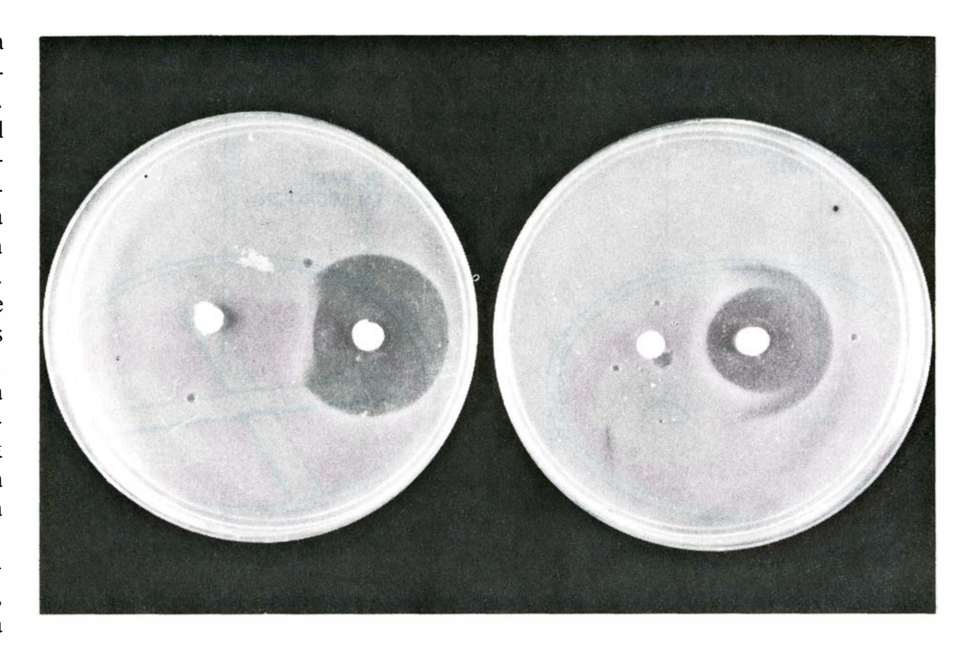
ESPECTRO DE ACCION ANTIBIOTICA para 25 microcinas, ensayadas sobre 60 tipos distintos de microorganismos pertenecientes a familias diversas. Los números entre paréntesis de la primera columna representan el número de organismos ensayados por cada género. Los resultados positivos se expresan con una intensidad de color proporcional a su potencia antibiótica. Las microcinas actúan preferentemente sobre bacterias muy parecidas a las cepas productoras (la mayoría *E. coli*), pero también inhiben el crecimiento de otros organismos, incluso tan distintos como las levaduras.

haber en el ecosistema intestinal? No es fácil dar una respuesta ni siquiera aproximada, ya que en nuestro caso el sistema de detección usado favoreció el aislamiento de ciertas enterobacterias aerobias y, además, como bacterias indicadoras de antibiosis solamente usamos cepas de *E. coli* de colección. Ensayos preliminares, utilizando como indicadoras cepas naturales, permiten sospechar que el número de microcinas detectables con estas estirpes puede ser muy superior. Con carácter indicativo, sí cabe decir que su número parece muy elevado. En nuestra experiencia concreta pensamos que alrededor de la mitad de las microcinas identificadas en nuestro estudio parecen distintas entre sí. En varios casos pudimos demostrar que una sola cepa era capaz de producir más de una microcina. Así, por ejemplo, cuando se trató de esclarecer la conducta anómala de una preparación del tipo antagonizable por metionina, se puso en evidencia la presencia simultánea de otra microcina no-antagonizable, causante de las observaciones contradictorias previas. Esta última, una vez separada de la microcina-met, resultó ser un oligopéptido pequeño e inestable que, como vimos más tarde, acompaña también a otras microcinas-met detectadas en cepas distintas.

Es interesante resaltar que las cepas productoras de microcinas han sido clasificadas hasta el presente como pertenecientes a la familia de las Enterobacteriaceae, dominando entre ellas claramente la especie *Escherichia coli* (83 por ciento) y en mucha menor proporción otras especies de los géneros *Klebsiella, Proteus y Enterobacter.* Por las razones aducidas más arriba, no podemos excluir la posibilidad de que especies de otras familias presentes en el mismo ecosistema intestinal u otros sistemas sean también productoras de antibióticos similares o parecidos a las microcinas; en este sentido resta por explorar el complejo mundo de las bacterias anaerobias estrictas de la flora intestinal, que pueden llegar a constituir el 90 por ciento de la población global de la misma.

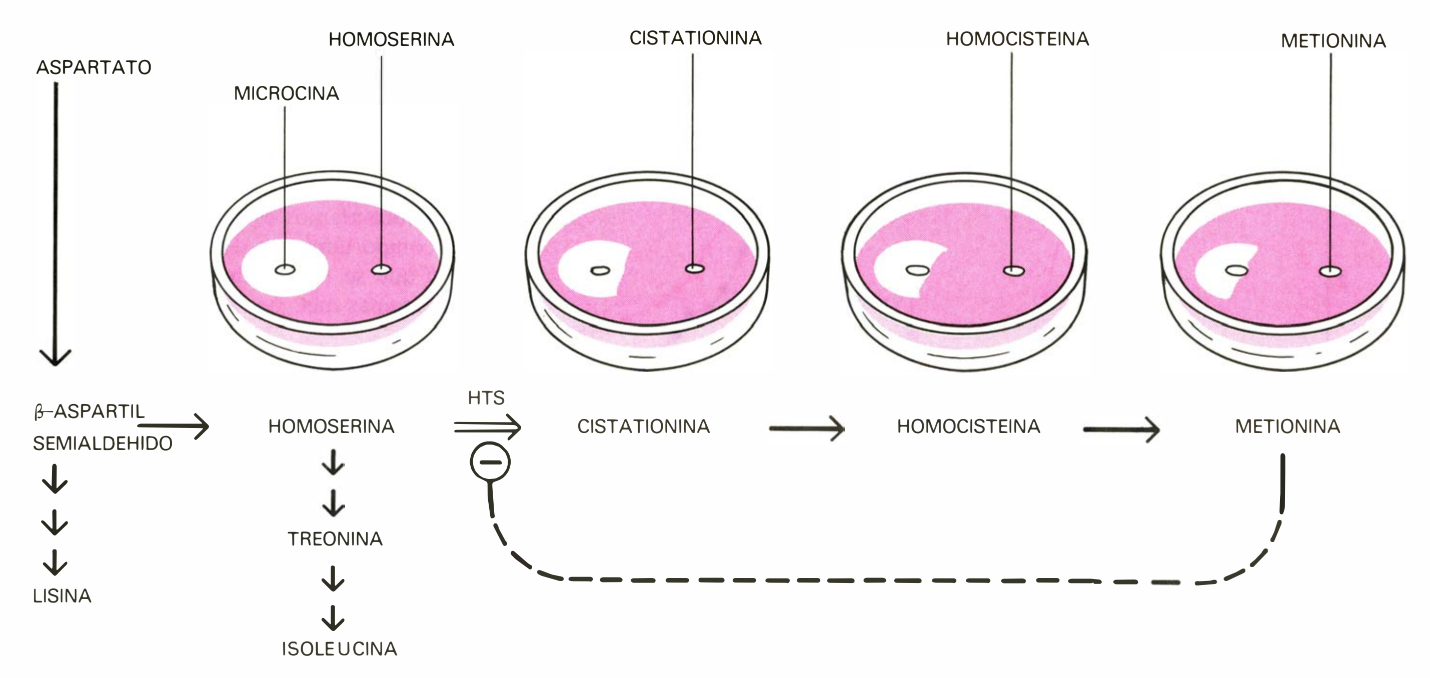
Uno de los aspectos más interesantes desde un punto de vista básico, y también aplicado, en el estudio de los antibióticos, se refiere a sus mecanismos de acción. En relación con las microcinas hemos empezado a abordar este problema particularmente con aquellas que hemos conseguido aislar con un elevado grado de pureza. Los resultados obtenidos muestran que los mecanismos implicados son de varios tipos. En unos casos actúan como bactericidas (que matan las bacterias sensibles), por ejemplo las microcinas 17 y 140, mientras que en otros, caso de las microcinas-met, actúan como bacteriostáticos (impiden simplemente el crecimiento), al menos a las dosis ensayadas. Estos modos de acción implican, a su vez, distintos mecanismos a nivel molecular, y así hemos encontrado que pueden afectar,

según los casos, la permeabilidad de la membrana, o la generación de bioenergía, o la síntesis de proteínas, entre otros. Hemos dedicado particular atención al mecanismo de acción de las microcinas-met, pues supusimos que la neutralización de su efecto antibiótico por la *l*-metionina habría de facilitarnos la exploración de sus bases bioquímicas. En este caso obtuvimos evidencia de que este tipo de microcinas inhibe la síntesis de proteínas de las células sensibles. Concomitante con esto encontramos una marcada inhibición en extractos acelulares del enzima denominado met-ARNt sintetasa, que es el responsable de una etapa clave para la incorporación de la metionina en la síntesis de las proteínas. Este efecto probó ser altamente específico y reversible por este aminoácido, pero las concentraciones necesarias para obtener una inhibición significativa resultaron ser bastante mayores que las requeridas para inhibir el crecimiento de las células enteras. Por ello pensamos que estas microcinas quizás afectasen de manera más crítica a algún otro nivel enzimático.

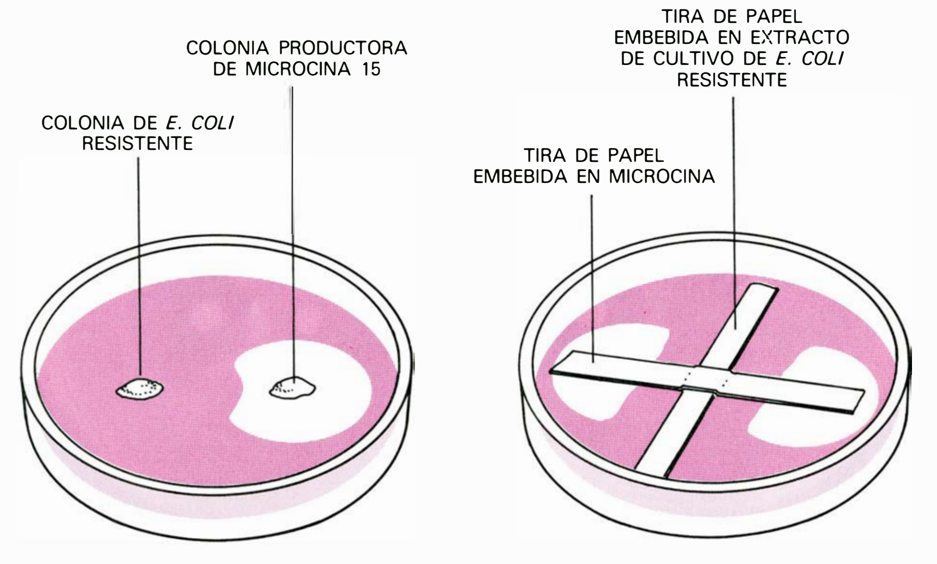


DETECCION DE UNA SEGUNDA MICROCINA producida por la cepa *E. coli* 15. En la placa de cultivo de la izquierda se colocaron dos pequeños discos de papel de filtro: en el del lado derecho se depositó una muestra de microcina 15; en el de la izquierda, una alicuota de una solución *l*-metionina. La fotografía muestra cierto antagonismo entre ésta y el efecto antibiótico. La placa de la derecha es igual que la anterior salvo que los discos de papel se colocaron más próximos, lo que permite evidenciar claramente la presencia de dos tipos de antibiosis, una de ellas inafectada por la metionina.

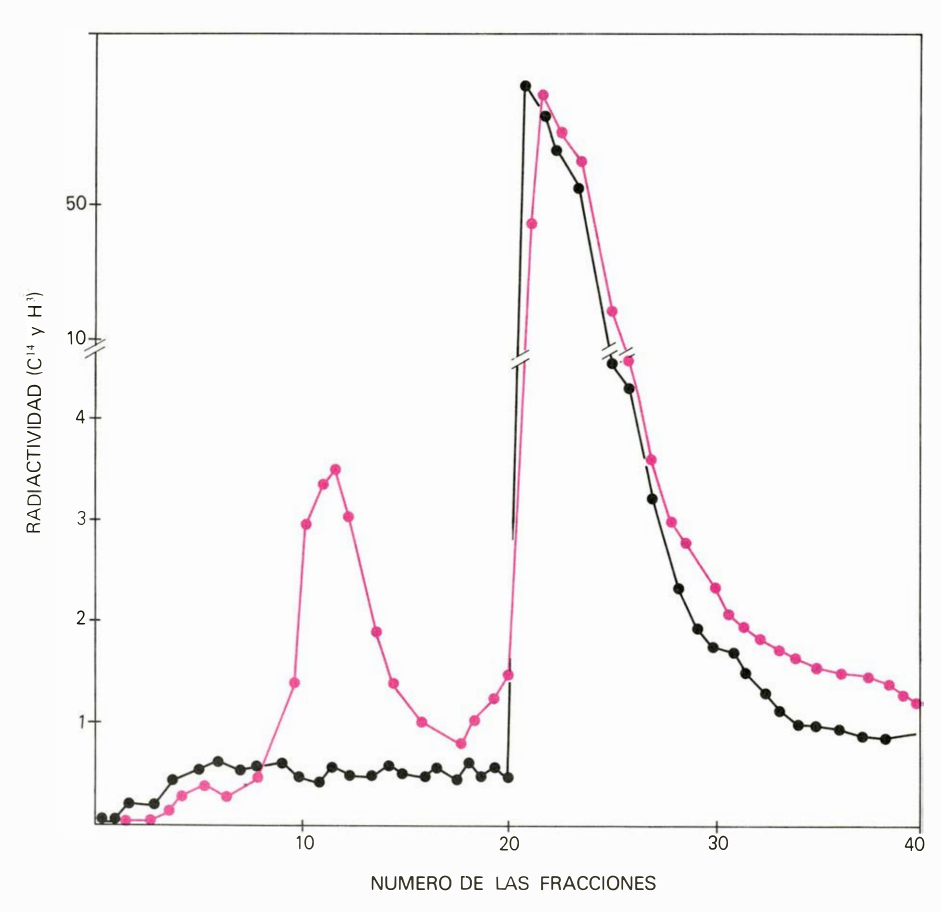
Efectivamente, éste parece ser el caso. La evidencia actualmente disponible sugiere de forma convincente que el efecto más notorio de las microcinas-met ocurre precisamente a nivel del enzima homoserin-O-transuccinilasa (HTS), que es el primero de la ruta metabólica que, partiendo del metabolito intermediario homoserina, conduce a través de varias etapas enzimáticas hasta la síntesis de metionina. El compuesto HTS es un enzima del tipo llamado alostérico, en el sentido de que es inhibido específicamente a ciertas concentraciones por el producto final de la ruta que con él se inicia, en este caso la propia metionina (y también por un derivado inmediato de ésta, la S-adenosilmetionina). Mecanismos como éste son muy útiles para la economía celular, pues autorregulan la síntesis de metabolitos esenciales para la célula, como son los aminoácidos, evitando así una hiper-(o hipo) producción innecesaria de estos compuestos por el simple expediente de controlar la primera etapa de su propia biosíntesis. En un experimento en que se utilizaron



BIOSINTESIS DEL AMINOACIDO METIONINA en la bacteria *E. coli*, según un esquema simplificado. En los discos situados a la izquierda de cada placa se ha depositado una muestra de microcina 15 purificada; en los de la derecha una cierta cantidad de uno de los metabolitos de la ruta de síntesis de la metionina: justamente el que se corresponde en esta ilustración debajo de cada placa. Las placas se habían inoculado previamente con una cepa de bacteria *E. coli* sensible a la microcina 15. Tras incubar por unas horas a 37 °C, en la que contiene metionina (extremo derecha) se ha antagonizado el efecto antibiótico de la microcina. Lo mismo ocurre con los precursores cistationina y homocisteína, pero no con homoserina. La hipótesis de los autores es que el efecto antibiótico principal de la microcina 15 se debe a que bloquea el enzima homoserin-O-transuccinilasa (HTS).



MECANISMOS DE RESISTENCIA a las microcinas-met, antagonizables por *l*-metionina. Tales mecanismos parecen implicar una pérdida de la autorregulación enzimática mediado por este aminoácido. En la figura de la izquierda se observa que una colonia productora de microcina 15 (tipo met) ejerce antibiosis a su alrededor, pero este efecto es antagonizado en la proximidad de otra colonia de *E. coli* resistente natural a esa microcina. Este fenómeno puede reproducirse *in vitro* utilizando extractos de cultivo de ambos tipos de células embebidos en tiras de papel cruzadas (ilustración de la derecha). En éste, como en otros casos, la sustancia responsable del antagonismo mostró ser *l*-metionina.



DISTINCION ENTRE ADN CROMOSOMICO Y PLASMIDICO, que puede conseguirse mediante la centrifugación de los lisados bacterianos en un gradiente especial de densidad. En esta gráfica se compara el ADN de una bacteria productora de la microcina 17 con el de una mutante derivada de la anterior que ha perdido la capacidad para sintetizar la microcina. El pico situado a la izquierda de la figura (*línea roja*) corresponde al ADN de un plásmido presente en la bacteria productora de microcina. Dado que éste desaparece en la mutante (*línea negra*) puede presumirse que dicho plásmido contiene información para la síntesis de la microcina 17. Los dos picos coincidentes a la derecha corresponden al ADN cromosómico de ambas cepas. El ADN en las distintas fracciones del grandiente se midió por su diferente radiactividad; en un caso debida a su contenido de carbono-14 y, en el otro, de tritio-(H^3).

células de *Escherichia coli* sensibles a la microcina 15 pudo observarse que los metabolitos cistationina y homocisteína, precursores de la biosíntesis de metionina, actúan como esta misma antagonizando el efecto de las microcinas-met, pero no así el primer precursor de esta vía, la homoserina.

Estos resultados son compatibles con la hipótesis formulada, la cual fue posteriormente apoyada por el estudio bioquímico de ciertas cepas resistentes a la acción de estas microcinas, según veremos más adelante, y por la prueba más directa obtenida *in vitro* utilizando una preparación acelular de HTS, cuya actividad catalítica es inhibida por la acción de la microcina 15 de forma similar a la ejercida por metionina.

¿Cómo podría explicarse la acción de las microcinas-met sobre la HTS? Todas las pruebas disponibles apuntan a que su acción inhibidora sobre este enzima ocurre por una suplantación o mimetización del efecto inhibidor alostérico de la propia *l*-metionina. El resultado de este efecto es el bloqueo de la síntesis de dicho aminoácido, lo cual determina la paralización de la biosíntesis de las proteínas celulares, ya que aquél es un componente imprescindible en la constitución de éstas. Lógicamente, la adición de *l*-metionina, u otro precursor inmediato de ésta, sortea dicho bloqueo y permite contrarrestar el efecto antibiótico.

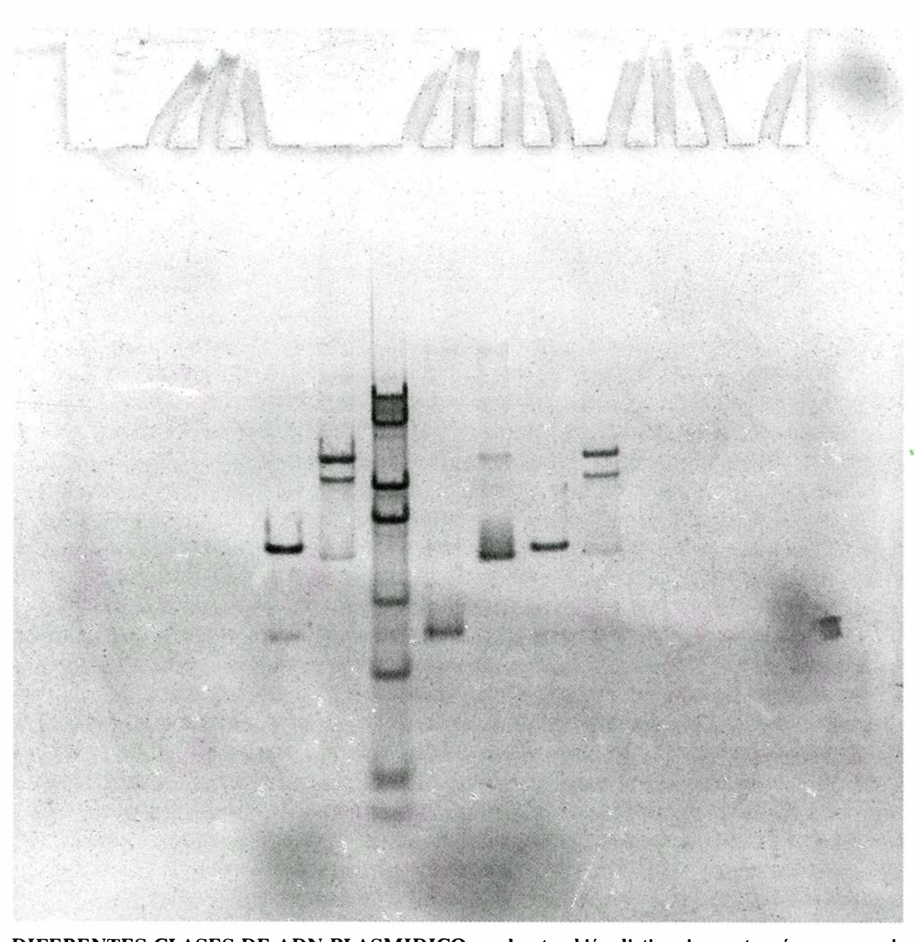
Los estudios actualmente en curso sobre la estructura molecular de las microcinas-met sugieren una gran analogía con la del aminoácido, lo que nos explica de un lado esta capacidad suplantadora y por otro, la prueba obtenida de que la *l*-metionina es un precursor esencial para la biosíntesis de aquéllas por las células productoras. Por añadidura, como ahora veremos, la hipótesis que formulamos es muy coherente con el comportamiento de las células de *E. coli* que se han hecho resistentes a las microcinas-met.

La adquisición de una capacidad genéticamente transferible para resistir la acción de un antibiótico dado, por ciertos microbios que antes eran sensibles al mismo, constituye un fenómeno bien conocido y relativamente frecuente, que polariza la atención de patólogos, microbiólogos, bioquímicos y otros especialistas relacionados con el tema. Sabemos que los mecanismos de resistencia pueden ser de varios tipos. Los experimentos que realizamos en el caso de cepas naturales o de mutantes de *E. coli* que se hicieron resistentes a la microcina 15, sugieren que la mutación lesionó

algún mecanismo regulador de la biosíntesis de metionina, probablemente el sitio alostérico al que se une este aminoácido en la HTS, de forma que la célula, al perder este mecanismo de control, por un lado produce un exceso de metionina, que aparece en concentraciones relativamente elevadas en el medio de cultivo, y, por otro, se hace insensible también al efecto inhibidor de la microcina. De ahí su resistencia al antibiótico.

Desconocemos de momento los mecanismos implicados en las mutaciones de resistencia a otros tipos de microcinas. En cualquier caso, estas mutantes son particularmente útiles para diferenciar e identificar aquellas entre sí, sobre la base de probar si la resistencia a una microcina dada implica también resistencia a alguna de las otras. La cuantía de esta resistencia cruzada indica en principio el grado de analogía entre aquellas.

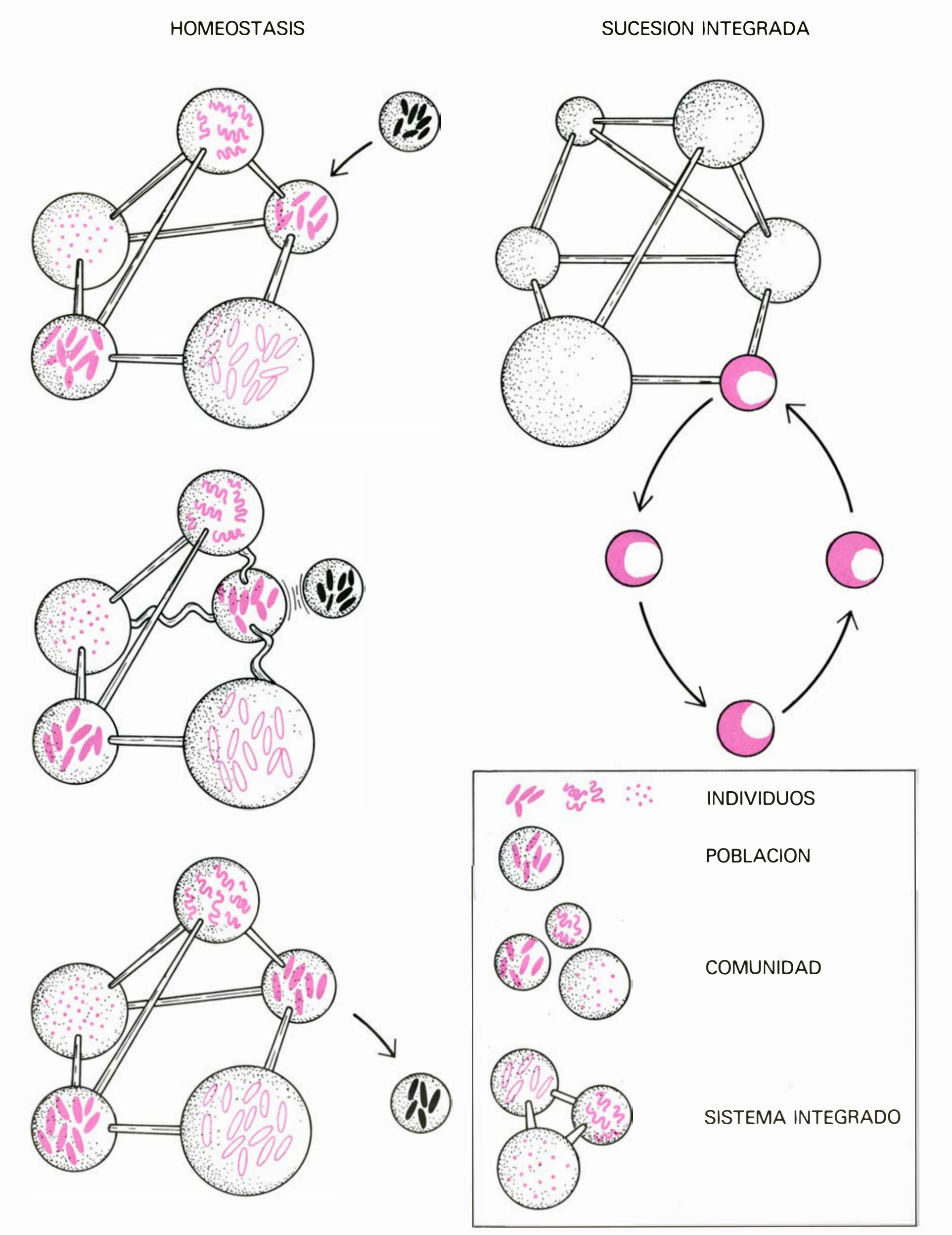
Una de las contribuciones más relevantes de nuestro trabajo fue la de constatar el carácter plasmídico, es decir, extracromosómico, del material genético codificador de las microcinas. Las primeras observaciones de esta línea fueron pronto confirmadas en el Servicio de Bacteriología Médica del Instituto Pasteur de París, en donde Daniel Bouanchaud pudo además aislar físicamente el ADN plasmídico de la microcina 17, que resultó tener una copia por célula, con un peso molecular de 36×10^6. Más recientemente, Royston, J. Claudio Pérez-Díaz y C. Clowes, de la Universidad de Texas en Dallas, han conseguido aislar el ADN plasmídico correspondiente a las microcinas-met 15, 93 y 136. En los tres casos mostró tener el mismo tamaño, con un peso molecular aproximadamente diez veces menor que el plásmido 17, si bien las bacterias productoras contienen unas 20 copias del correspondiente plásmido por célula. Resultados similares fueron obtenidos por Bouanchaud con el plásmido 93.



DIFERENTES CLASES DE ADN PLASMIDICO pueden también distinguirse entre sí por su movilidad electroforética en un gel de agarosa. En las columnas *C* y *D* pueden verse las bandas, respectivamente, de un plásmido de resistencia (plásmido *R*) y de un plásmido que codifica para la producción de las microcinas 15*a* y 15*n* (plásmido *M*). Estas estructuras de ADN pueden ser troceadas por el uso de enzimas de restricción que atacan secuencias específicas; en *A* y *F*, se observan los tres fragmentos que el enzima Eco *R* 1 ha producido en el plásmido *R*; en *B* y *G*, los dos fragmentos que el enzima Bam *H* 1 provoca en el plásmido *M*. En *E* se observan los fragmentos que Eco *R* 1 produce en el ADN del fago *P* 22 que se utiliza como referencia (J. C. Pérez-Díaz y R. C. Clowes, de la Universidad de Texas).

La disponibilidad de ADN microcínico purificado facilitará sin duda el estudio de las distintas microcinas a través de los potentes recursos de la ingeniería genética. Uno de los problemas más intrigantes pendiente de esclarecer se refiere a las bases moleculares y a los mecanismos factuales de la "autoinmunidad" de las células a sus propias microcinas. ¿Cómo es, por ejemplo, que la cepa *E. coli* K 12, sensible a las microcinas-met, se vuelve resistente al recibir por conjugación sexual los plásmidos de estas microcinas desde las cepas productoras originales?

Tras la descripción que hemos hecho de los rasgos más sobresalientes de las microcinas, podemos volver ahora a plantearnos la hipótesis original que determinó precisamente su búsqueda. Parece evidente que las propiedades de estos agentes antibióticos son razonablemente coherentes con las idealmente concebidas al comienzo de nuestro trabajo. Incluso su especificidad de acción y su carácter plasmídico nos ofrecen un cuadro sugerente de implicaciones ecológicas en consonancia con lo que sabemos sobre la flora intestinal. Así, la participación de las microcinas en este ecosistema podría explicarnos, por ejemplo, las sucesiones normales y altamente específicas entre bacterias coliformes, o la rapidez y contagiosidad hospitalaria, fenómeno frecuente en clínicas infantiles, que muestran ciertas cepas enteropatógenas, que posiblemente son capaces de colonizar y de desplazar a otras enterobacterias mediante el concurso de plásmidos muy infectivos codificadores de potentes microcinas. A su vez, la eficaz homeostasis del ecosistema intestinal podría explicarse en buena parte por la capacidad de ciertas microcinas para impedir el crecimiento de otros microorganismos alejados filogenéticamente de las cepas productoras; un caso concreto seria el de la levadura patógena *Candida albicans*, residente inerte y adventicio en las mucosas bucales, pero que es capaz de invadir masivamente nuestro tracto digestivo cuando destruimos la flora autóctona por la aplicación de ciertos antibióticos convencionales. Lo mismo podríamos decir de otros procesos, hoy relativamente frecuentes, debidos a infecciones por *Pseudomonas, Proteus, Staphylococcus, Serratia,* etcétera. Desde un punto de vista evolutivo, seria difícil explicar la abundancia de cepas productoras de microcinas sin asignar a éstas una misión en este sentido. Ciertamente, un planteamiento ecológico de este tipo sólo podría aceptarse explícitamente con el apoyo de pruebas directas. En relación con esto, algunos experimentos en marcha son alentadores, particularmente la observación de un notorio pa-



ORGANIZACION de las poblaciones de microorganismos en el intestino. Dicha organización cumple con las condiciones de un sistema integrado y, por ello, rechaza el asentamiento de otros microbios extraños al ecosistema (homeostasis), pero permite, sin embargo, la sustitución recurrente de ciertas poblaciones por otras parecidas constituidas por individuos de la misma especie (sucesión integrada). La especificidad antibiótica de las microcinas es coherente con su posible protagonismo en ambos fenómenos, y también como agente de control del tamaño de las poblaciones que integran el sistema.

ralelismo entre las características de la actividad antibiótica detectada en extractos de algunas heces infantiles (sujetos no tratados por antibióticos) y las de las microcinas producidas en cultivos de laboratorio por aislados entéricos de esas mismas heces. Ultimamente, también hemos evidenciado varios casos en los que los microorganismos aerobios que colonizan inicialmente el contenido intestinal del recién nacido, disponen de microcinas activas sobre otras bacterias colonizadoras aisladas previamente y, luego, desaparecidas en los mismos sujetos.

Independientemente de un enfoque ecológico, las microcinas constituyen por varias razones un nuevo grupo de agentes biológicos distintos de los antibióticos conocidos. Estos proceden de microorganismos esporogénicos y aeróbicos cuyo hábitat normal es el suelo; las microcinas de bacterias anaerobias facultativas no-esporogénicas, vinculadas al ecosistema intestinal de animales superiores. Los antibióticos convencionales se sintetizan por los organismos productores tras la fase de crecimiento exponencial, en la llamada fase estacionaria o idiofase, en donde ocurren ciertos procesos de diferenciación celular (como es la esporogénesis). Por el contrario, las microcinas se producen y excretan al medio durante el período de desarrollo exponencial.

El origen y distribución de los antibióticos, al nivel actual de conocimiento, confrontan al microbiólogo con varios interrogantes pendientes de resolver. Es bien sabido el hecho de que el 80 por ciento de los antibióticos conocidos son sintetizados por un solo tipo de bacterias filamentosas, pertenecientes al género *Streptomyces.* Los restantes se distribuyen entre unos pocos géneros más, bien de la familia a la que pertenecen los anteriores (Actinomicetales) o del género *Bacillus,* o de ciertos tipos de hongos (entre ellos la penicilina). ¿Cuál es la singularidad biológica que confiere a los estreptomicetos, y en menor grado a otros grupos, esa extraordinaria "fantasía" en la síntesis química, que por otra parte implica una correspondiente riqueza en su información genética? El contenido en ADN y el grado de mutabilidad espontánea de estos organismos no parecen justificar su extraordinario potencial de capacidades biosintéticas. La explicación habría que buscarla posiblemente en las características de su hábitat natural, el suelo, en donde cabe un número virtualmente infinito de nichos distintos, a los que verosimilmente se han ido adaptando por vía evolutiva los estreptomicetos y otros microorganismos asociados con ellos. La presencia de una constelación de antibióticos tan rica en formulaciones estructurales sería así un reflejo de esta variación ecológica.

Este plantemiento, sin embargo, no nos explica por qué otras especies colonizadoras habituales del suelo no producen aparentemente antibióticos. ¿Podría ser que, en realidad, sí que los producen, pero que aún no han sido detectados? El caso de las microcinas parece apoyar esta última hipótesis. Siendo la bacteria *Escherichia coli* probablemente la especie biológica más explorada y mejor conocida desde un punto de vista bioquímico y genético, se nos revela ahora como productora de una gama insospechada de entidades antibióticas. ¿Cabría pensar que esta capacidad está también presente en otros microorganismos? Así lo creemos nosotros. Por otra parte, no es necesario invocar el argumento de la variabilidad ecológica para asegurar una gran versatilidad genética en el modelo del colibacilo, ya que a éste le cabe disfrutar de un vasto reservorio genómico por la vía plasmídica. Los plásmidos constituyen paquetes de información codificada en ácidos nucleicos que, por el hecho de su transferabilidad de unas células a otras, facilitan enormemente la

variabilidad biológica sin comprometer los límites físicos impuestos a cada célula para acomodar y manipular su propio ADN. De esta forma *E. coli* y otros microorganimos pueden adquirir un gran caudal de expresión genética para la síntesis de microcinas o de otros agentes biológicos de potencial interés adaptativo.

Será necesario valorar en el próximo futuro cuál podría ser la utilidad de las microcinas en terapeútica. Un espectro antibiótico interesante y diversificado, un bajo peso molecular y presumiblemente una escasa o nula toxicidad (se están produciendo probablemente de forma "normal" en nuestro intestino) apoyan sus posibilidades a este respecto. Sin embargo, su eventual utilización terapéutica requerirá primero una ampliación de nuestro conocimiento básico sobre estos agentes, desde varios puntos de vista, incluyendo el estudio de sus estructuras moleculares. Tanto este abordaje como la prevista investigación farmacológica en animales de laboratorio, requerirán la obtención de cantidades apreciables de microcinas en estado puro, que de momento presenta bastantes dificultades, debido a la muy baja concentración de estos agentes en los caldos de cultivo originales. La ayuda que nos prestan los laboratorios Antibióticos, S.A. y Normón, S.A., con la aportación de equipos pilotos de fermentación y otras facilidades, acelerará sin duda el desarrollo de la investigación básica sobre microcinas.

En cualquier caso, el plantemiento original que nos llevó al descubrimiento de estos agentes estuvo presidido por una clara orientación ecológica, que entendemos debería seguir informando la futura exploración farmacológica. A lo largo de estas líneas hemos esbozado nuestra creencia de que, en el proceso evolutivo del hombre, se ha llegado a establecer una convivencia con ciertos ecosistemas bacterianos que, en rigor, han coevolucionado con nosotros mismos, de suerte que algunos de sus mecanismos homeostáticos constituyen hoy nuestros propios guardianes. El gran microbiólogo Selman A. Waksman, refiriéndose a los antibióticos convencionales, sentenció, con resonancia bíblica, que: "Tú salvación vendrá de la tierra". Quizás el estudio de los mecanismos de protección y agresión que frente al invasor microbiano exógeno presentan los microorganismos del intestino humano, permita el descubrimiento de medios más "fisiológicos" en la lucha antiinfecciosa, liberando, una vez más, al hombre de la dependencia de la tierra.

Juegos matemáticos

Problemas ajedrecísticos en otro plano, incluidos giros y simetrías

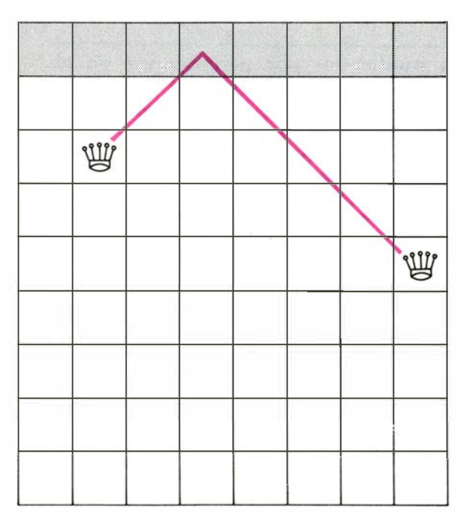
Martin Gardner

"Es un hermoso arte, complejo y estéril, tan sólo relacionado con la forma ordinaria del juego en la medida en que, pongamos por caso, puedan aplicar propiedades de la esfera tanto un malabarista que prepara un nuevo número como un jugador de tenis que se esfuerza en vencer un torneo. La mayoría de jugadores de ajedrez... acusan solamente un tibio interés por estas charadas extravagantes y mudables con la moda."

—VLADIMIR NABOKOV, *Speak, Memory*

La última ocasión en que dediqué al ajedrez una sección completa en *Scientific American* fue hace siete años (mayo de 1972). Este mes el ajedrez vuelve a ser tema, bajo la forma de ciertos insólitos problemas que no habían sido analizados hasta ahora. Al hablar de "problemas" no me estoy refiriendo a los del tipo habitual, que piden el número de jugadas para que alguno de los jugadores dé jaque-mate. Las tareas propuestas son de naturaleza más matemática; tanto así que para trabajar en ellas basta que los lectores conozcan las reglas del juego.

Comenzaré por un notable descubrimiento debido a David A. Klarner, matemático de la Universidad estatal de Nueva York en Binghampton. A principios de la década de los sesenta, los especialistas en teoría de números estaban trabajando en un problema relativo a cualquier conjunto de enteros consecutivos, de 1 a $2n$, siendo n cualquier entero positivo. Un tal conjunto puede siempre descomponerse en dos subconjuntos de igual tamaño: uno, el formado por los enteros consecutivos de 1 a n, otro, el formado por los enteros sucesivos, desde $n + 1$ hasta $2n$. El problema consiste en averiguar si para distintos valores de n es posible emparejar cada número de uno de los conjuntos con un número del otro, de forma que las $2n$ sumas y diferencias absolutas (donde se prescinde del signo de la diferencia) sean todas distintas. Por ejemplo, cuando $n = 1$ el conjunto original es $\{1, 2\}$ y sus subconjuntos son $\{1\}$ y $\{2\}$. En este caso el único emparejamiento posible da una solución trivial, porque la suma $1 + 2$ es igual a 3, y distinta del valor absoluto de la diferencia 1 - 2, que es igual a 1.



Reinas que se atacan por reflexión

El perspicaz descubrimiento de Klarner fue reconocer que el problema anterior es isomorfo a una variante del antiguo problema ajedrecístico de situar n reinas sobre un tablero de lado n de forma que ningún par de reinas se ataquen. Su variante consistió en colocar una fila más en lo alto del tablero, que él denomina banda de reflexión, porque, como se muestra en línea de color en la ilustración de esta página, las reinas pueden atacarse por reflexión en la banda. Dicho de otra forma, se admite que las reinas pueden atacarse entrando diagonalmente en la banda y saliendo por la diagonal opuesta. En su artículo titulado "The Problem of Reflecting Queens" (*The American Mathematical Monthly*, vol. 74. n.º 8, págs. 953-955; octubre de 1967) Klarner demostró que el problema de teoría de números antes citado admite solución si y sólo si es posible situar n reinas sobre un tablero de n por n provisto de banda reflectante sin que ningún par de ellas se ataque, ni directamente, ni por reflexión en la banda.

La ilustración superior de la página siguiente muestra cuán fácilmente se transforman las soluciones de esta variante del problema de colocación de reinas (en el caso presentado, con $n = 4$) en soluciones del problema numérico. Comenzando por el ángulo superior izquierdo del tablero se numeran consecutivamente las filas, desde arriba hacia abajo, continuando después la sucesión de izquierda a derecha, al pie de cada columna. Obsérvese que las dos soluciones presentadas (que en este caso son las únicas) son imágenes simétricas izquierda-derecha, como si cada una de ellas viera a la otra en un espejo. Se hace ahora una lista de los números de fila y columna —coordenadas— de todas las reinas, como se muestra bajo los tableros de la ilustración. Calculando las sumas y diferencias de los ocho pares de números obtenidos, es inmediato comprobar que se dispone de una solución del problema aritmético en el caso $n = 4$.

No es difícil descubrir la razón de que ambos problemas sean isomorfos. Es evidente que el conjunto de números escritos a un lado y debajo del tablero es precisamente el conjunto de números de 1 hasta $2n$, y que el conjunto de índices de filas de las damas de una solución forma parte de la primera mitad de este conjunto, mientras que los índices de columna forman parte de la otra mitad. Si hubiera un par de reinas que se amenazasen por una fila o columna, uno de los conjuntos de índices contendría un número repetido. Al ser imposible que así ocurra, los subconjuntos de índices han de estar formados por los números de 1 a n, y de $n + 1$ hasta $2n$, respectivamente. Además, si dos reinas se amenazasen por una diagonal ascendente hacia la derecha, se repetiría el valor absoluto de alguna diferencia, mientras que si dos reinas se amenazasen por reflexión en la banda, habría una suma y una diferencia absoluta que serían iguales. (De amenazarse por diagonal descendente, se repetirían dos sumas.) No obstante, dado que ningún par de reinas puede atacarse mutuamente, ni por reflexión ni directamente, las sumas y las diferencias absolutas de los pares de números deducidos de la solución ajedrecística han de ser todas ellas distintas, con lo que proporcionan una solución del problema aritmético.

Se conocen desde antiguo soluciones de la versión clásica del problema de las reinas (sin banda reflectante) para tableros de hasta octavo orden; pueden consultarse en el capítulo 16 de mi libro *The Unexpected Hanging and Other Mathematical Diversions* (Simon & Schuster, Inc., 1972). No existe solución para los

tableros de orden 2 y 3; si se desprecian las soluciones deducidas de otras por giros o simetrías, para el orden cuatro hay exactamente una solución "fundamental". Hay dos soluciones fundamentales en orden cinco, una en el orden seis, seis de orden siete y 12 de orden ocho. Para obtener todas las soluciones del correspondiente problema aritmético, a partir de uno de estos conjuntos de soluciones, es necesario tener en cuenta los cuatro posibles giros de cada solución fundamental, así como las deducidas por simetría axial izquierda-derecha, y descartar todos los casos donde las reinas puedan amenazarse por "rebote" en una banda adosada al borde superior de cada tablero. Los casos que todavía subsistan dan un sistema de soluciones del problema aritmético.

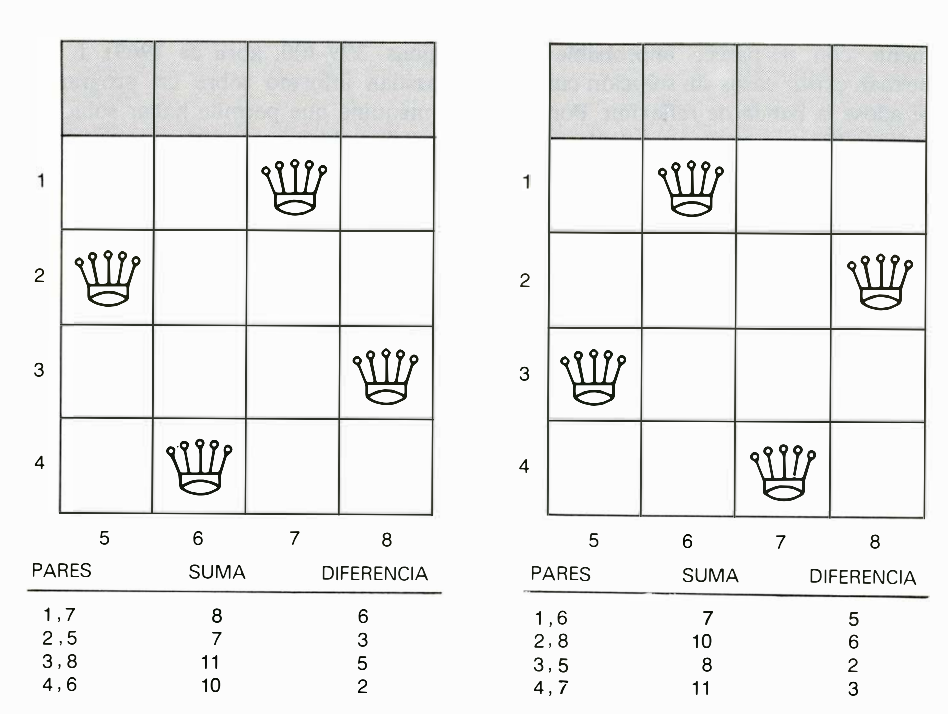
En el caso $n = 4$ la disposición de las piezas es invariante por rotaciones, por lo que no es preciso ensayar las deducidas por giro. Para $n = 5$ las cinco damas pueden situarse de dos formas esencialmente diferentes, que se muestran en la ilustración inferior de esta misma página. En la disposición izquierda se observa una amenaza por reflexión (que se indica en línea de color). Esta disposición es invariante por rotaciones. Por consiguiente, ninguna de las configuraciones deducidas de ellas mediante giros o simetrías puede engendrar soluciones del problema aritmético. La segunda disposición genera una solución orientada como en la figura, y otra más haciéndola girar 90 grados en el sentido de las manecillas del reloj. Las imágenes deducidas de éstas por simetría axial dan otras dos soluciones más, lo que da un total de cuatro.

La única configuración de orden 6 cuyas reinas no se amenazan contiene ataques por reflexión en todas las colocaciones deducidas de ella por giro o simetría axial, así que el correspondiente problema aritmético carece de solución. Por consiguiente, tampoco podrá resolverse el siguiente "solitario" con naipes. Se colocan en fila cartas de valor 1, 2, 3, 4, 5 y 6. El objetivo del problema consiste en situar debajo de cada una de estas cartas otra de valor 7, 8, 9, 10, sota o reina (se supone que la sota y la reina tienen valores 11 y 12, respectivamente) de forma que los 12 números obtenidos al tomar la suma y la diferencia (en valor absoluto) del par de cartas de cada columna sean todos distintos.

En el caso $n = 7$ tan sólo uno de los giros de tan sólo una de las seis soluciones básicas genera solución del problema numérico. La configuración mencionada más la imagen deducida por simetría engendran, en total, dos soluciones. De las 12 soluciones fundamentales del problema de reinas no amenazadas del tablero ordinario de orden ocho, tres poseen un giro que las depura de amenazas por reflexión, y hay todavía otra configuración que admite dos de tales giros. Estas cinco disposiciones, más sus imágenes simétricas originan 10 soluciones distintas del problema numérico en el caso $n = 8$. Los lectores podrán hallarlas todas ensayando los giros y simetrías de las 12 soluciones fundamentales que se muestran en la página 192 de *The Unexpected Hanging.*

Para tableros cuadrados de orden superior al octavo es mucho más sencillo preparar programas de ordenador que calculen el número total de soluciones (incluidas todas las simetrías y giros) de la variante del problema de las reinas, que escribir el programa que expurgue de los giros y simetrías de las soluciones del problema clásico las que contengan ataques por reflexión. No obstante, algunos programas recientes han ampliado el número de soluciones fundamentales para los órdenes de 9 a 16. Los resultados obtenidos son, respectivamente, 46, 92, 341, 1787, 9233, 45.752, 285.053 y 1.846.955. Los números correspondientes a los órdenes 9, 10 y 11 fueron correctamente determinados antes de generalizarse el empleo de computadores, pero los resultados para órdenes superiores contradicen todos los asertos y conjeturas que conocía por libros más antiguos.

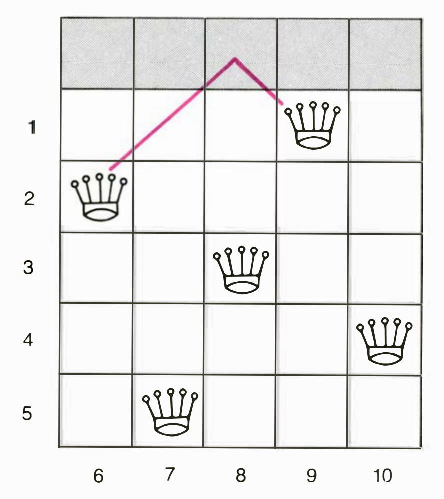
En vista de que el número de solucio-

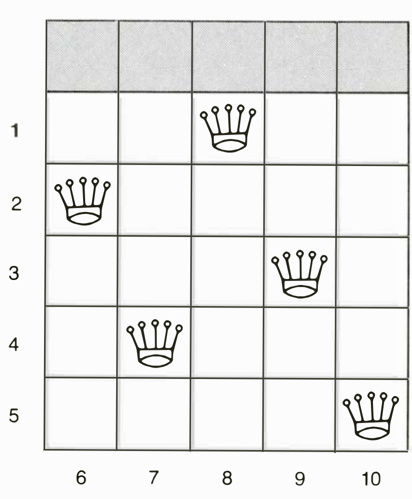


PARES	SUMA	DIFERENCIA
1,7	8	6
2,5	7	3
3,8	11	5
4,6	10	2

PARES	SUMA	DIFERENCIA
1,6	7	5
2,8	10	6
3,5	8	2
4,7	11	3

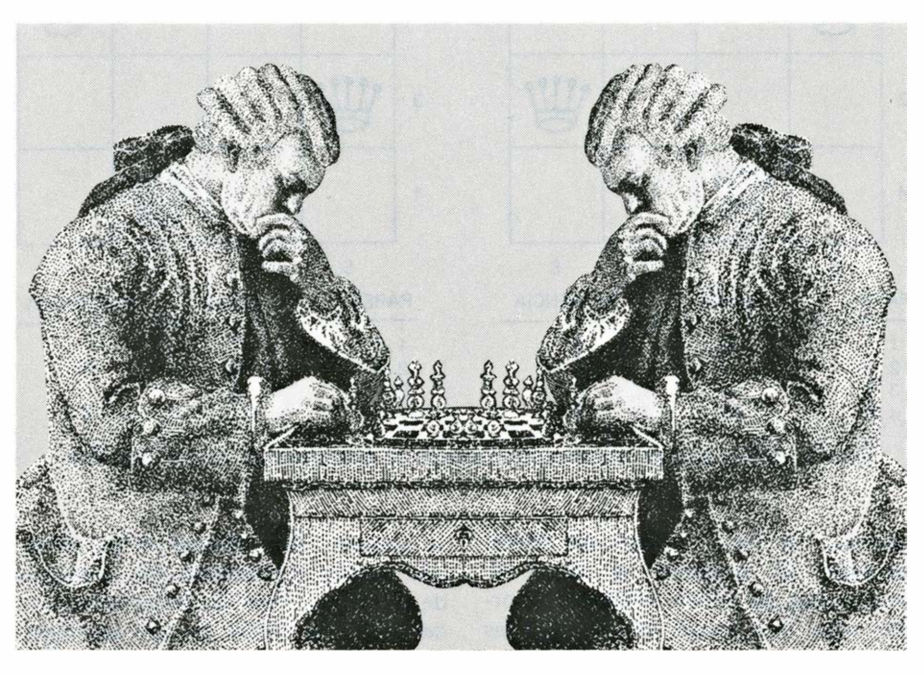
Cuatro reinas que no pueden amenazarse ni directamente ni por reflexión



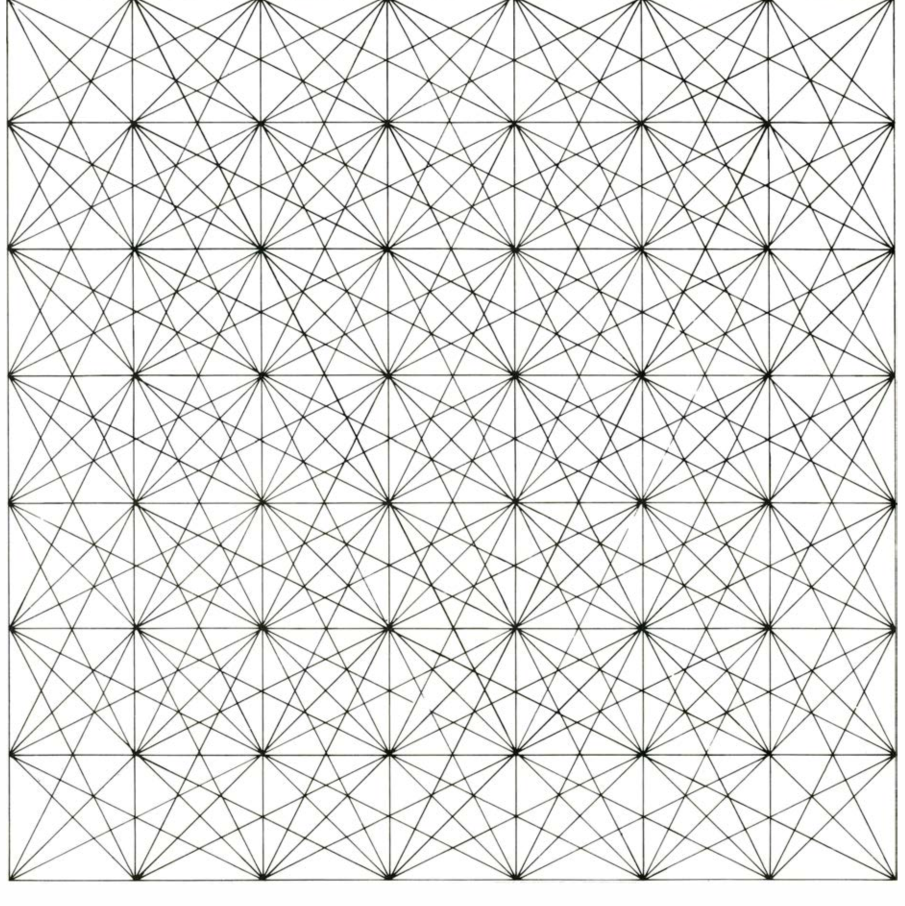


Problema clásico de reinas no amenazadas. En el caso **n** *= 5 tiene dos soluciones fundamentales*

nes fundamentales crece muy rápidamente con *n*, parece improbable que puedan existir casos sin solución cuando se adose la banda de reflexión. Por este motivo Klarner ha enunciado la conjetura de que para todo *n* mayor que 6 el problema aritmético tiene cuando menos una solución. En *The American Mathematical Monthly* (vol. 76, n.º 4, páginas. 399-400; abril de 1969), J. D. Sebastian informó sobre un programa de máquina que permite hallar soluciones del problema numérico para todos los valores de *n* comprendidos entre 9 y 27. Sin embargo, todavía se ignora el número de soluciones distintas que posee el problema cuando *n* es un número mayor que 8.



Ajedrez simétrico de Mitsumasa Anno



Las 1840 diferentes jugadas de un juego de ajedrez

La asimetría de la posición inicial de las piezas de ajedrez debida a la diferencia entre dama y rey hace que las piezas blancas sean imagen "especular" (simétrica) de las negras. La ilustración superior de esta misma página muestra la forma en que el dibujante japonés Mitsumasa Anno ha extendido la simetría de las piezas a los propios jugadores. En razón del apasionado interés de Anno por temas de inspiración matemática, se le ha apodado "el Escher japonés". La ilustración se ha tomado de su libro *Anno 1968-1977*.

La figura de Anno hace pensar en dos retorcidos rompecabezas de ajedrez publicados en 1866 por Sam Loyd, máximo creador americano de problemas y rompecabezas ajedrecísticos. En ambos problemas se exige que cada movimiento del Negro haya de ser copia por simetría de la última jugada del Blanco. (Evidentemente, estamos interesados en las jugadas posibles, lícitas y no en jugar competentemente.) El primer problema consiste en descubrir una partida "simétrica" en donde las blancas den mate en su cuarta jugada. El segundo –muchísimo más difícil– es inventar una partida simétrica en la que el Blanco hace una octava jugada que obliga al Negro a darle mate con una jugada no simétrica. (Los estudiosos de problemas de ajedrez llaman "suimates" a las jugadas de este tipo.) El mes que viene describiré partidas de ambos tipos que, aunque extravagantes, son perfectamente lícitas.

¿Cuántas jugadas diferentes pueden hacerse en un tablero ordinario de ajedrez? La respuesta es 1840. Cada posible movimiento está representado por un trazo rectilíneo en el grafo que se muestra en la ilustración inferior de esta misma página. Este grafo se ha tomado de la tarjeta de felicitación navideña, para el año 1972, del estadístico I. J. Good. Se han sustituido las 64 casillas del tablero por 64 puntos, con lo que el grafo de Good parece mostrar un tablero de 7 por 7.

El grafo de Good hace pensar en un problema no trivial de coloreado de grafos, que ha sido aplicado a tableros cuadrados. Dada una cualquiera de las cinco piezas distintas de ajedrez, ¿cuál es el mínimo número de colores necesarios para colorear las casillas de un tablero, de forma que dondequiera que se sitúe la pieza tan sólo pueda moverse a casillas de distinto color? El número mínimo anterior se llama número cromático del tablero de orden *n* correspondiente a la pieza.

El mínimo número cromático es 2, que es el número de colores precisos

para un caballo en todo tablero de orden mayor o igual que tres. El coloreado que resuelve el problema en cualquiera de estos tableros consiste simplemente en pintarlos escaqueados. Tal descubrimiento resuelve el siguiente rompecabezas: Si cada casilla de un tablero cuadriculado está ocupada por un caballo, ¿podrán todos ellos moverse al mismo tiempo a distinta casilla? El problema tiene solución si y sólo si el lado del tablero tiene longitud par, porque únicamente entonces hay iguales números de casillas de cada color. Cuando el lado es de longitud impar hay una casilla más de uno de los colores. En tal caso, como cada caballo ha de saltar a casilla de distinto color quedará por lo menos un caballo sin sitio adonde ir.

El número cromático del rey en todo tablero de orden dos o mayor es cuatro. Asignándoles a los colores las letras *A, B, C* y *D*, la fila superior del tablero está coloreada *ABABAB...;* la fila segunda, *CDCDCD...;* la tercera, *ABABAB...,* y así sucesivamente. Para una torre, el número cromático de un tablero de orden *n* es *n*. Para colorear el tablero se comienza en un ángulo, pintando a partir de él líneas diagonales paralelas, usando para la única casilla de la primera diagonal el color *A*, el color *B* para las dos casillas de la segunda diagonal, el *C* para las tres de la tercera, y así sucesivamente, hasta usar el *n*-ésimo color para la diagonal principal. Volviendo a empezar por *A* para la diagonal siguiente, se usa la misma sucesión de colores para las diagonales menores.

El número cromático de un alfil, en un tablero de orden *n* es *n* o *n*-1, según que la máxima diagonal que pueda recorrer lícitamente esté formada por *n* casillas o por *n*-1. Si el alfil puede recorrer *n* cuadros, se pintan todas las filas de distinto color; en cambio, si el máximo recorrido es de *n*-1 casillas, el color de la primera fila puede volver a utilizarse para la última (la enésima).

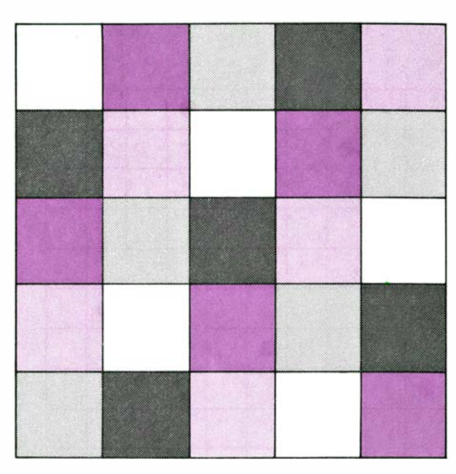
El problema tiene máximo interés en el caso de la reina, porque entonces equivale al antiguo problema de forma "cuadrados latinos diagonales", esto es, matrices *n* por *n* cuyas casillas se han pintado de forma que en ninguna fila o columna ni diagonal –cualquiera que sea su longitud– se repite un mismo color. El número mínimo de colores necesarios será, evidentemente, igual al número cromático de la reina en el tablero de orden *n*. Hace más de 60 años que George Pólya demostró que con *n* colores pueden construirse dameros *n* por *n* de este tipo si y solamente si *n* no es divisible ni por 2 ni por 3. El método de Pólya puede verse en la ilustración de esta página. Obsérvese que cada color define líneas paralelas de saltos de caballo. (Estas cadenas son todavía más evidentes en la configuración correspondiente al tablero de 11 por 11, tomada de la página 252 de la reimpresión publicada por Dover de las *Mathematical Recreations* de Maurice Kraitchik.) Mediante pautas semejantes pueden engendrarse soluciones para todo tablero de orden *n*, siempre que *n* no sea múltiplo de 2 ni de 3.

Fijémonos en lo que sucedería en cualquier tablero así cuando en todos los cuadros rojos se colocasen reinas rojas, reinas azules en todos los cuadros azules, reinas verdes en todos los cuadros verdes, y así sucesivamente. Como es obvio, el tablero de *n* por *n* permitiría situar en él *n* conjuntos de *n* damas del mismo color, de forma que ninguna de ellas amenace a otra de su mismo equipo. El mínimo tablero de este tipo (aparte el de 1 por 1, que es trivial) es de orden cinco. Ha sido vendido comercialmente muchas veces en forma de rompecabezas formado por 25 fichas divididas en cinco conjuntos de cinco de cada color. El problema consiste en situar las fichas de forma que ninguna de ellas comparta con otra de su color una misma fila, columna o diagonal.

Que yo tenga noticia, todavía no ha podido hallarse el número cromático de la reina en todos los tableros de orden *n* cuando *n* es múltiplo de 2 o 3, si bien se conoce la solución para muchos tableros de orden no muy alto. Por ejemplo, el tablero de seis por seis requiere siete colores, y el de ocho por ocho precisa nueve. Sí ha podido demostrarse, en cambio, que el número cromático de la reina no puede ser menor que *n* ni tampoco mayor que $n + 3$.

El método de coloreado de Pólya tiene un atributo adicional, que algunos llaman propiedad envolvente: al unir un par cualquiera de lados opuestos del tablero coloreado resulta un cilindro en el que continúa verificándose que toda "diagonal" contiene *n* colores. Con terminología algo distinta, cada "diagonal partida" del cuadrado contiene los *n* colores. Los cuadrados que gozan de esta propiedad se llaman cuadrados latinos pandiagonales.

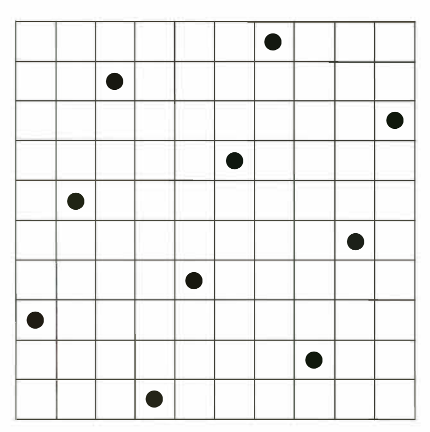
La amazona, o superdama, es una pieza especial que reúne en sí los movimientos de la dama y del caballo de ajedrez. Cuando *n* es menor que 10, es imposible situar sobre un tablero de orden *n* otras tantas superdamas sin que algunas de ellas se amenacen. La única solución en el caso $n = 10$ se muestra en la ilustración superior de la página siguiente. Solomon W. Golomb ha demostrado la existencia efectiva de tales configuraciones cuando *n* sea mayor


Un cuadrado latino pandiagonal de orden cinco

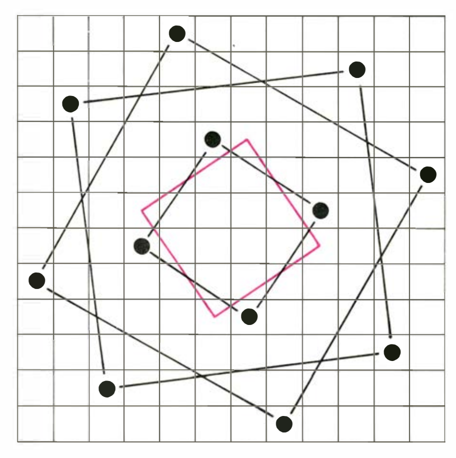
que 9 y número primo o una unidad menor que un número primo. Ashok K. Chandra ha demostrado que existen configuraciones envolventes de *n* superdamas no amenazadas, si y sólo si *n* es mayor que 11 y no es múltiplo de 2 ni 3. Por mi parte, desconozco si existe solución general al problema de las amazonas.

Sin Hitotumatu, de Kyoto, preparó un programa de ordenador que produjo seis soluciones para el tablero de orden 11, si bien las seis resultan ser idénticas si el tablero se arrolla en ambos sentidos, construyéndose un toro. Estas soluciones fueron ya anteriormente descubiertas por Golomb, y utilizan el método de construcción de éste. Para encontrar una de ellas basta ampliar la solución de orden 10 que se muestra en la ilustración superior de la página siguiente, añadiéndole una fila en lo más alto y una columna a su lado derecho; la undécima superdama se coloca entonces en el ángulo superior derecho del tablero, como se muestra en color. Hitotumatu encontró también 23 soluciones fundamentales para el problema de las superdamas en el tablero de orden 12. La presentada en la segunda ilustración de la página siguiente es probablemente la más bella, por la colocación simétrica de dos cuadrados iguales y del cuadro interior pequeño. (Las líneas de color muestran cómo el cuadrado pequeño puede modificarse mediante giros.)

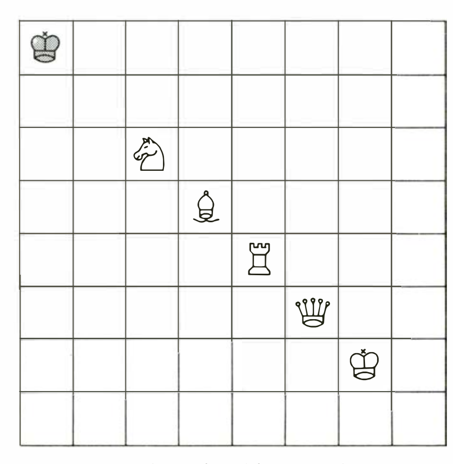
Donald E. Knuth ha descubierto recientemente un método para resolver los casos de órdenes 14 y 15, y que puede generalizarse a ciertos órdenes superiores. Se sabe ahora que combinando las construcciones de Knuth y de Chandra, existen soluciones al problema de las amazonas siempre que *n* sea mayor que 9 y no sea de la forma $12k + 8$ o $12k + 9$ para algún entero *k*. El método de Chandra da soluciones para los casos $n = 6k$, $n = 6k + 1$, $n = 6k + 4$ y



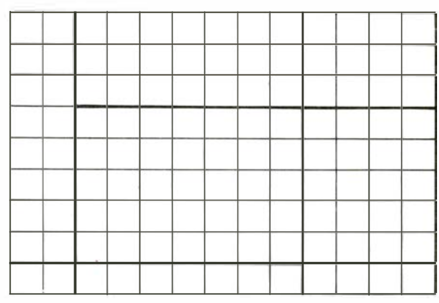
Tablero mínimo de **n**-*por*-**n** *para* **n** *superdamas*



Doce superdamas en un tablero de 12-por-12



¿Cómo podrán las blancas dar mate en menos de una jugada?



Un rectángulo rectangulado

$n = 6k + 5$; el método de Knuth da soluciones para la mitad de los casos restantes. Por consiguiente, los casos mínimos no resueltos son $n = 20$, $n = 21$, $n = 32$ y $n = 33$.

Concluiré con una pregunta: ¿Cómo conseguirá el Blanco, en la posición mostrada en la tercera de las ilustraciones, empezando por arriba, de esta misma página, dar jaque mate al Negro, en una fracción de jugada?

El primer problema del mes pasado se refería al juego de apuestas con naipes que su creador, Robert Connelly, bautizó "Diga Roja". En ese juego el banquero va dando cartas boca arriba, que extrae de una baraja ordinaria (francesa). En algún momento, antes de que saque la última carta, el jugador debe decir: "Roja". Si la próxima carta que salga es roja, el jugador gana la partida; en caso contrario pierde. El jugador tiene derecho a inspeccionar las cartas ya extraídas en todo momento. Es posible utilizar esta información para diseñar una estrategia que eleve por encima de 1/2 su probabilidad de ganar en una larga serie de partidas.

Aunque parezca sorprendente, no puede. Pueden darse muchas demostraciones formales de este hecho, pero quizá lo mejor sea "ver" lo que sucede analizando el juego cuando hay dos cartas negras y dos cartas rojas. Si llamamos *N* a las negras y *R* a las rojas, hay seis ordenaciones que son igualmente probables:

RRNN
RNRN
RNNR
NNRR
NRNR
NRRN

Ojeando esta columna es evidente que la probabilidad de que gane el jugador es 1/2, tanto si pide roja antes de la primera, de la segunda o de la tercera carta. Por otra parte, supongamos que se hayan sacado ya las dos cartas negras. Se podría razonar que la tercera será forzosamente roja, con lo que el jugador no puede perder. Es cierto; pero hay que tener en cuenta que, a la larga, tal eventualidad solamente se presentará en uno de cada seis juegos. En los otros cinco, la probabilidad de que el jugador venza es de 5/6 x 2/5, o sea, 1/3, con lo que la probabilidad total de ganar en la tercera carta es de 1/6 + 1/3 = 1/2.

Probemos ahora una idea distinta. Supongamos que la estrategia del jugador sea cantar "roja" si, al sacar la primera carta, ésta es negra. De no ser así, esperará hasta que falte solamente una carta por sacar. Si la primera carta es negra, el jugador tiene probabilidad 2/3 de ganar en la siguiente; pero esta ventaja está compensada porque si la primera carta no es negra, tiene probabilidad 2/3 de perder en la última carta.

El razonamiento anterior se generaliza sin dificultad a barajas de *n* cartas negras y *r* cartas rojas. La probabilidad de que un jugador llegue a vencer usando la estrategia que desee es siempre $r/(n + r)$. Así pues, el conocimiento de las cartas del mazo ya extraídas en nada favorece al jugador, sin que importen ni la relación de cartas negras a rojas, ni el tamaño del mazo. Ninguna estrategia es superior a la de pedir roja en la primera carta, o en la última, o en realidad, en cualquier otro momento. (Como ya expliqué el mes pasado, el jugador debe hacer su apuesta antes de empezar el juego; de lo contrario podría aprovecharse mediante un sistema como el del blackjack, apostando fuerte cuando tenga la suerte de cara, y flojo cuando sus probabilidades sean pequeñas.)

El segundo problema consistía en diseñar un conjunto de cuatro cartas que permita un truco de magia parecido al descrito el mes pasado, aunque ahora los cuatro números aparentemente elegidos al azar han de tener siempre el mismo producto, en vez de tener suma constante. El secreto consiste en elegir los pares de números a situar en las caras de cada tarjeta de forma que siempre tengan la misma razón. Por ejemplo, en las cartas preparadas por Shigeo Futigawa, el ilusionista japonés que inventó ambas variantes del juego, los pares de números son 26/34, 39/51, 52/68 y 65/85. En todos los casos, la razón es de 13 a 17.

El truco se desarrolla casi exactamente igual que el descrito el mes pasado, aunque ahora los números seleccionados se multiplican en lugar de sumarse. En esta variante el producto predicho es de 5.860.920. Obsérvese bien que este número es igual a $13 \times 13 \times 17 \times 17 \times 2 \times 3 \times 4 \times 5$. Lo mismo que hice con la versión del mes pasado dejaré al cuidado de los lectores la demostración algebraica de que el truco no puede fallar.

En el número de mayo, el rectángulo-gura superior de la página 112 fue dibujado incorrectamente. El rectángulo de la parte inferior derecha del rectángulo grande no debe tener tres unidades de ancho, sino cuatro, como se ve en la ilustración inferior que aparece en esta misma página.

Taller y laboratorio

Experimentos con el método de Edwin Land para obtener color a partir del blanco y negro

Jearl Walker

En 1950 Edwin Land llevó a cabo una serie de experimentos simples con idea de investigar los estímulos responsables de la visión de los colores. En uno de ellos fotografió dos veces, en blanco y negro, una misma escena. Al hacer la primera fotografía colocó delante del objetivo un filtro rojo y al hacer la segunda un filtro verde. De los negativos obtenidos sacó dos diapositivas en blanco y negro, que proyectó con dos proyectores de forma simultánea sobre una misma pantalla, colocando delante de la diapositiva realizada a través del filtro rojo otro filtro del mismo color y no poniendo nada, es decir, proyectando sin filtro la otra diapositiva, la obtenida a través del filtro verde.

Lo que intuitivamente se podría esperar de tal proyección conjunta era una reproducción de la escena original en varios tonos de blanco, rosa y rojo, debido al filtro colocado en uno de los proyectores. Pero, sorprendentemente, Land vio sobre la pantalla una reproducción de la escena original con la mayoría de los colores que ésta tenía en la realidad. Aunque los colores aparecían un poco débiles, estaban allí a pesar de que lo que Land había proyectado eran dos diapositivas en blanco y negro.

Durante más de 20 años, Land, John L. McCann y sus colaboradores se han dedicado a descifrar cómo el sistema de visión humano (el ojo y la corteza visual) percibe y es capaz de diferenciar la sutil diferencia entre las diversas tonalidades de los colores. Land describió sus primeros trabajos en *Scientific American*, en mayo de 1959. Los experimentos realizados en los últimos veinte años le han permitido profundizar en el conocimiento de los procesos de la visión en color. Tales avances los explica Land en su último artículo publicado [véase "La teoría retinex de la visión del color", por Edwin H. Land, INVESTIGACION Y CIENCIA, febrero 1978].

Robert Szabo, alumno de la Universidad estatal de Cleveland, repitió no hace mucho los experimentos de Land: sacó diapositivas en blanco y negro a través de filtros de colores y proyectó luego simultáneamente las diapositivas sobre una pantalla, si bien utilizó un solo filtro. Aunque los resultados no dieron unos colores tan vivos como al parecer lo fueron los obtenidos en los primeros trabajos de Land, sí demostró cuán fácil resultaba para un fotógrafo aficionado obtener los efectos que conseguía Land.

Land fotografiaba unas veces naturaleza muerta, frutas por ejemplo, y objetos similares, y otras fotografiaba un papel coloreado con una complicada distribución de colores que él llamaba un "Mondrian". Szabo por su parte eligió como modelo un nuevo cartel, el ColorChecker, producido recientemente por Macbeth, una filial de la Kollmorgen Corporation. El cartel se está convirtiendo en una referencia típica en todas las ramas de la fotografía en color, gracias a la distribución espectral de sus cuadrados cromáticos. Cada cuadrado tiene un color casi idéntico al color natural asequible al fotógrafo. Así, un cuadrado tiene un azul muy similar al típico azul del cielo, pero no sólo similar en el tono de azul dominante sino también a lo largo de todo el espectro visible.

Una ventaja de fotografiar referencias standard es que uno puede comparar en cualquier momento el modelo con las imágenes proyectadas. Un modelo con una naturaleza muerta no tiene tal constancia cromática , sobre todo si en el modelo hay frutas. Otra ventaja es que los resultados que uno obtenga pueden transmitirse a otros con tal de que empleen el mismo modelo. La disposición de colores Mondrian no gozaba de esa flexibilidad.

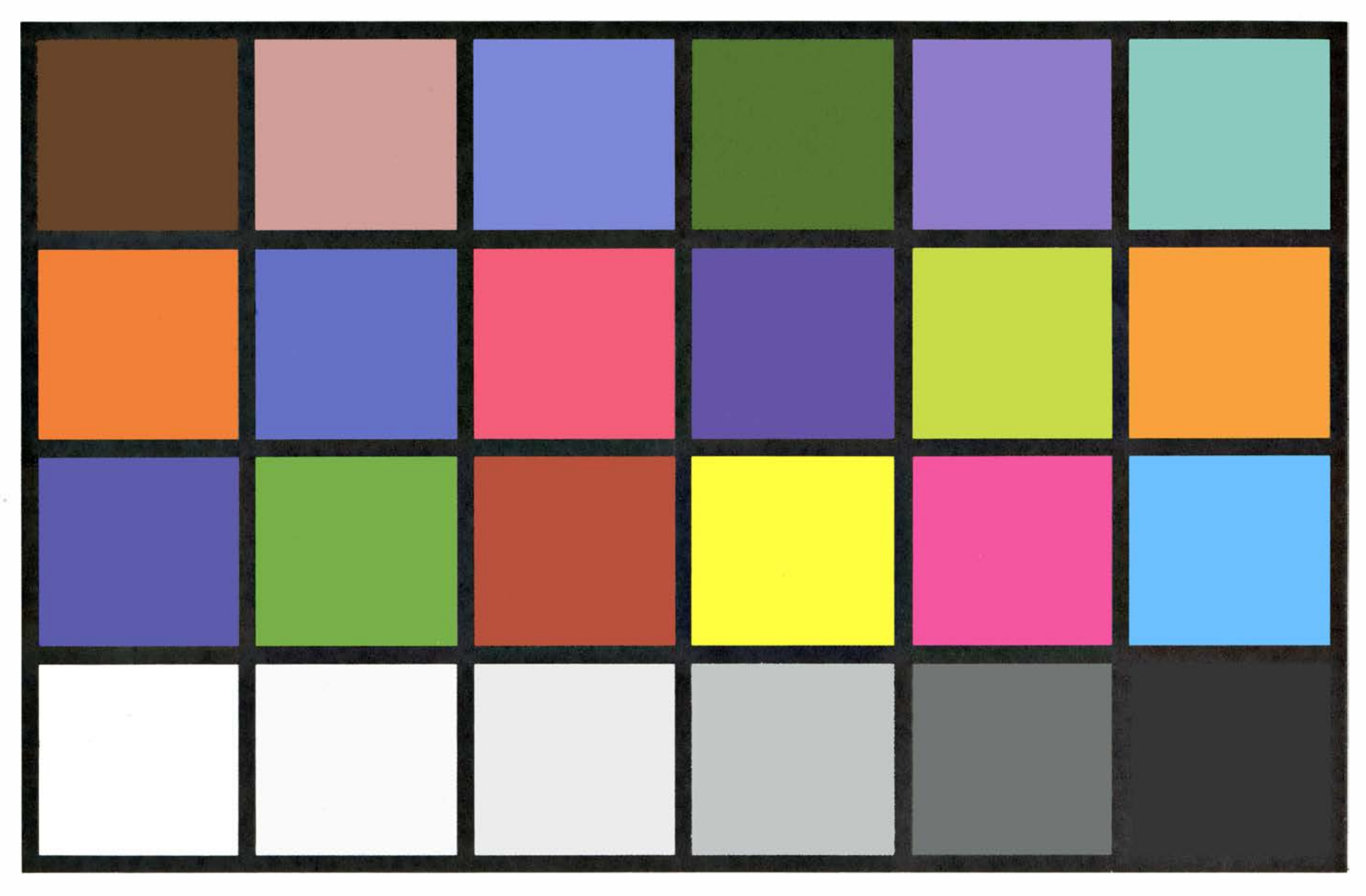
El ColorChecker tiene también una desventaja: la influencia de los cuadrados circundantes sobre cada uno de los cuadrados de la imagen proyectada puede afectar a los resultados del experimento. Con la disposición aleatoria de las pequeñas secciones de Mondrian, cada una de ellas formada por varios colores y colocada de un modo distinto, se anula la influencia de los colores circundantes.

Para fotografiar la tabla de los colores, Szabo montó su cámara en un soporte rígido colocándole un cable disparador con el fin de evitar que se moviera en la acción de dispararla. Iluminó la tabla cartel con dos reflectores de tungsteno orientados de manera que ambos formaran un ángulo de 45 grados con el panel donde estaba colocada la tabla. En vez de filtros de colores de alta calidad, Szabo empleó otros algo más baratos, pero suficientemente buenos. Los filtros medían 12,5 x 20 centímetros, un tamaño adecuado para la cámara fotográfica y para el proyector.

El mejor tipo de película a emplear quizá sea el de diapositivas en blanco y negro de Polaroid, ya que dichas películas son sensibles de un modo bastante uniforme en las distintas regiones del espectro visible. Szabo optó por las películas Pan-X para blanco y negro que son relativamente baratas. Con un equipo Kodak para revelar directamente al positivo, Szabo reveló y montó él mismo las diapositivas. Primero fotografiaba la tabla de colores ajustando la cámara para obtener la exposición (apertura de diafragma y velocidad de exposición) adecuada a las condiciones de iluminación que tenía, y luego realizaba otras tomas modificando dichos parámetros. Con el fin de comparar y ver posteriormente la influencia de los mismos, anotaba con sumo cuidado los datos de las condiciones en que había tomado cada foto. La proyección sobre una pantalla de tipo convencional la efectuaba con dos proyectores. Tanto la luz empleada en el proyector como la usada para realizar las fotos han de ser blancas y de buena calidad, de forma que ambas presenten distribuciones cromáticas muy similares.

A la diapositiva hecha con el filtro que básicamente deja pasar las longitudes de onda más largas se le llama registro largo, mientras que la otra, hecha con el filtro que deja pasar sobre todo las longitudes de ondas más cortas, recibe el nombre de registro corto. Siguiendo los experimentos de Land, Szabo hizo un gran número de ensayos empleando una amplia gama de filtros y de métodos de proyección.

El tipo más sencillo de ensayo era el que realizaba con un registro largo y uno corto; cuando empleaba un filtro en la proyección éste era el mismo filtro con el que había hecho el registro largo y lo situaba en el proyector que iba a proyectar dicha diapositiva. En uno de estos ensayos, hizo el registro largo con



La tabla ColorChecker fotografiada por Robert Szabo en sus experimentos

Marrón oscuro	Cuero oscuro	Azul celeste	Verde hoja	Lila	Verde azulado
Naranja	Azul púrpura	Rosa	Púrpura	Amarillo verdoso	Amarillo naranja
Azul	Verde	Rojo	Amarillo	Magenta	Cian
Blanco (0,05)†	Neutro 8* (0,23)†	Neutro 6.5* (0,44)†	Neutro 5* (0,70)†	Neutro (3,5)* (1,05)†	Negro (1,50)†

* Notación cromática † Densidad óptica

Identificación de los colores en la tabla

un filtro Edmund 821, que transmite principalmente la luz roja. El registro corto lo hizo con un filtro 856, que deja pasar el azul. Szabo había conseguido dos diapositivas en blanco y negro, hechas al fotografiar el mismo objeto dos veces, empleando sendos filtros que transmitían en extremos opuestos del espectro visible. Las diapositivas no mostraban, en apariencia, ninguna indicación sobre los colores y, caso de que llevaran información cromática de algún tipo, era de esperar que no incluyese ni el verde ni el amarillo, ya que ninguno de los dos filtros transmite bien las longitudes de onda correspondientes al verde y al amarillo.

Cuando Szabo proyectó las diapositivas, colocó frente a la diapositiva de registro largo el filtro 821. Al superponer las dos imágenes se encontró con una reproducción de la tabla ColorChecker que, además del rojo, mostraba algo de azul, amarillo y verde (pero no púrpura).

Las diapositivas que Szabo proyectó habían sido obtenidas fotografiando la tabla con el diafragma y la velocidad de exposición adecuados a la luz que suministraban los reflectores. Pero al proyectar las diapositivas realizadas con tiempos de exposición algo mayores, o algo menores, encontró por lo general que las características de color que obtenía eran similares a las anteriores, siempre que no se hubieran sobreexpuesto o infraexpuesto demasiado. Land había dicho que el nivel de intensidad de los proyectores no era crítico. Cualesquiera que fuesen los códigos de color, las diapositivas en blanco y negro mostraron que no dependían de forma crítica de la intensidad de la luz. Szabo y yo comprobamos esto, montando dos láminas polarizadoras de forma cruzada delante de uno de los proyectores. Modificando el ángulo que formaban los ejes de polarización de cada lámina, podíamos variar la intensidad de luz de la imagen proyectada en la pantalla. Vimos que dicha intensidad no influía críticamente en la proyección de los colores.

En otro ensayo, Szabo usó como registro corto una diapositiva hecha a través de un filtro 817 y como registro largo una ya hecha con el filtro 821. El filtro 817 deja pasar más el naranja que el 821. De nuevo, el filtro 821 se situó delante del proyector que proyectaba el registro largo. La superposición de imágenes dio una proyección muy débil y predominantemente rosada. Las características de transmisión de los filtros eran demasiado similares como para alcanzar todos los colores del espectro en la proyección. Si hubiésemos utilizado el mismo filtro al sacar las dos diapositivas, en la proyección posterior no hubiéramos conseguido más color que el del filtro situado delante de uno de los proyectores.

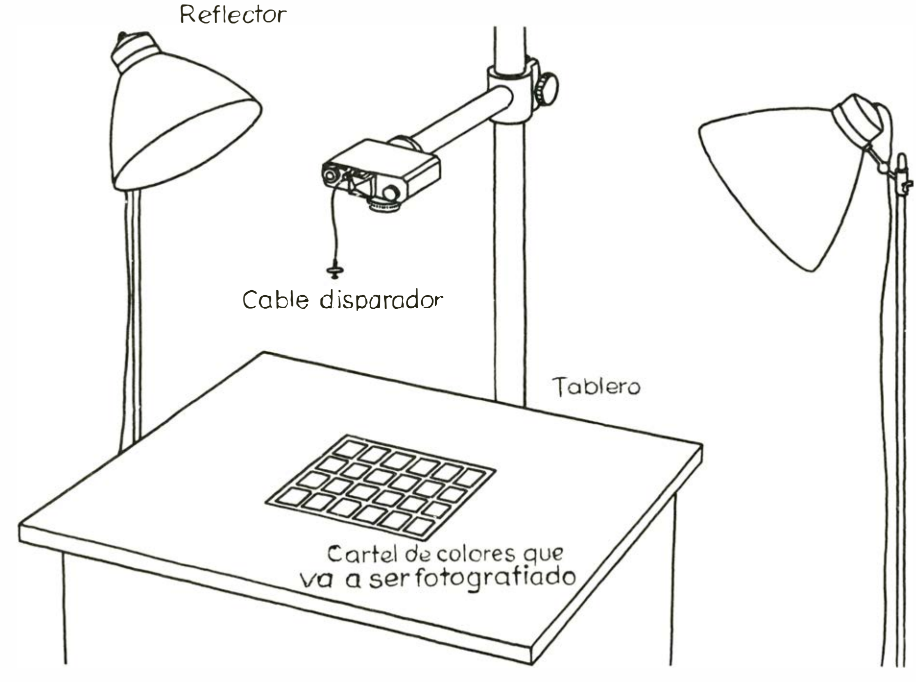
En el artículo publicado por Land en *Scientific American* se incluía un gráfico con el que se podía predecir qué colores no aparecerían en la proyección. Por ejemplo, Szabo proyectó sus diapositivas hechas a través de un filtro 821 (el cual presenta una transmisión máxima para las longitudes de ondas comprendidas entre 650 y 700 nanómetros) y de un filtro 856 (que tiene su máxima transmisión aproximadamente en 475 nanómetros). El punto correspondiente a dichos filtros en el gráfico está situado en una zona denominada "no púrpura". Szabo no veía el púrpura.

En la zona del gráfico a lo largo de la diagonal del mismo, el registro corto y el largo están demasiado próximos uno del otro como para producir colores en la imagen proyectada. Los filtros 817 y 821 empleados por Szabo corresponden a un punto próximo a la diagonal. Una zona sin delimitar debajo de la diagonal corresponde al caso de que el filtro empleado para hacer la diapositiva de registro largo se utilice luego para proyectar el registro corto, y viceversa. A Szabo y a mí nos divertía mantener el filtro del registro largo delante del proyector de registro largo y luego situarlo delante del proyector del registro corto para obtener la inversión de los colores.

Szabo ensayó la proyección de diapositivas anteponiendo al proyector del registro corto el filtro del registro corto y no poniendo ningún filtro delante del proyector de registro largo. No se lograron colores tan bien definidos. Por ejemplo, realizó fotografías usando filtros 858 (que transmite el verde) y 863 (que transmite más en la zona del azul). En la imagen proyectada se apreciaban marrones, azules y azules verdosos.

Szabo experimentó también con el incremento del número de filtros, es decir: usó dos o más filtros a la vez con el fin de estrechar la banda de longitud de ondas transmitida. En un ensayo superpuso los filtros 861, 863 y 866 para eliminar cualquier luz fuera del rango de 550 a 700 nanómetros, que corresponde al amarillo, rojo y algo del verde. Hizo la diapositiva correspondiente al registro largo con un filtro 821 y la proyectó con el mismo filtro situado delante de su proyector. No encontró el amarillo en la proyección, pero en cambio sí estaban los otros colores.

En una serie de experimentos, Szabo fotografió el registro largo a través de dos filtros (el 874 y el 877) superpuestos, para conseguir una banda estrecha de paso situada alrededor de los 500 nanómetros. El registro corto lo hizo a través de un filtro 866, el cual transmite mucho el azul, con un máximo de transmisión en los 450 nanómetros. Cuando situó los filtros utilizados para hacer el registro largo delante del proyector del registro largo, obtuvo azules y verdes. Por el contrario, si colocaba el filtro correspondiente al registro corto delante del registro largo, obtenía azules y un amarillo pálido (esto es, una inversión de colores, dado que delante del registro largo se había colocado el filtro "erró-



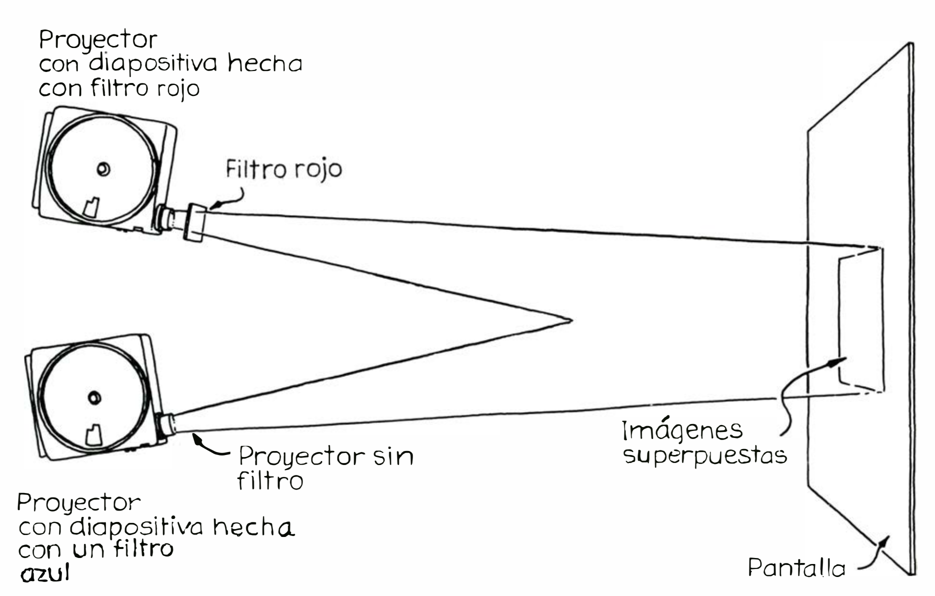
Montaje de Szabo para fotografiar la tabla de colores

neo"). Llevado por la curiosidad, juntó los filtros y colocó delante de la diapositiva, registro largo, otro filtro, el 807, que deja pasar un ancho de banda que va de los 525 a los 775 nanómetros. Aunque este filtro no era el adecuado para la proyección, algunos colores débiles se vieron en la pantalla.

Land exponía en el artículo mencionado que los colores salían aun cuando el registro corto se hubiera fotografiado sin filtro. Szabo combinó algunas diapositivas hechas sin filtro y proyectadas sin filtro con algunos de los registros largos anteriormente obtenidos. Al colocar frente al registro largo su correspondiente filtro, aparecieron en la pantalla colores débiles.

¿Por qué aparecen los colores cuando se proyectan a la vez sobre una pantalla diversas diapositivas? No se sabe aún la razón detallada de este fenómeno, aunque Land y sus colaboradores han avanzado bastante en el tema durante las dos últimas décadas. Al observar una escena con varios colores, éstos no se ven cada uno por una determinada zona sino que, de acuerdo con la teoría de los colores, en la retina humana hay tres pigmentos visuales que actúan como tres fotómetros independientes. Puesto que las curvas de absorción de los pigmentos tienen sus respectivos máximos a 440, 535 y 565 nanómetros, cada uno de los pigmentos es sensible en una determinada zona del espectro y entre los tres abarcan el espectro visible completo. Esta teoría mantiene que, cuando la energía que llega a la retina correspondiente a una longitud de onda de, por ejemplo, 565 nanómetros, es mayor que la de los otros dos máximos, el observador percibirá el área referida en un tono rojizo. Pero si la energía llega a 440 nanómetros, se percibirá de color azul.

Esta era la explicación dada en los libros de texto durante mucho tiempo, por lo menos en las explicaciones sucintas de la teoría del color. Land ha demostrado, sin embargo, que el color es independiente del flujo energético que llega al ojo, cualquiera que sea su longitud de onda. Así, a lo largo de un experimento provocó que la energía procedente de dos zonas distintas fuera la misma en los tres máximos de absorción. De acuerdo con la teoría del color, ambas zonas deberían verse del mismo color; ahora bien, como las dos zonas eran dos partes distintas del cartel llamado Mondrian, el observador vio una zona verde y otra roja. En consecuencia, los flujos energéticos por sí sólos no eran la fuente de la información que sobre los colores recibía y procesaba el observador.

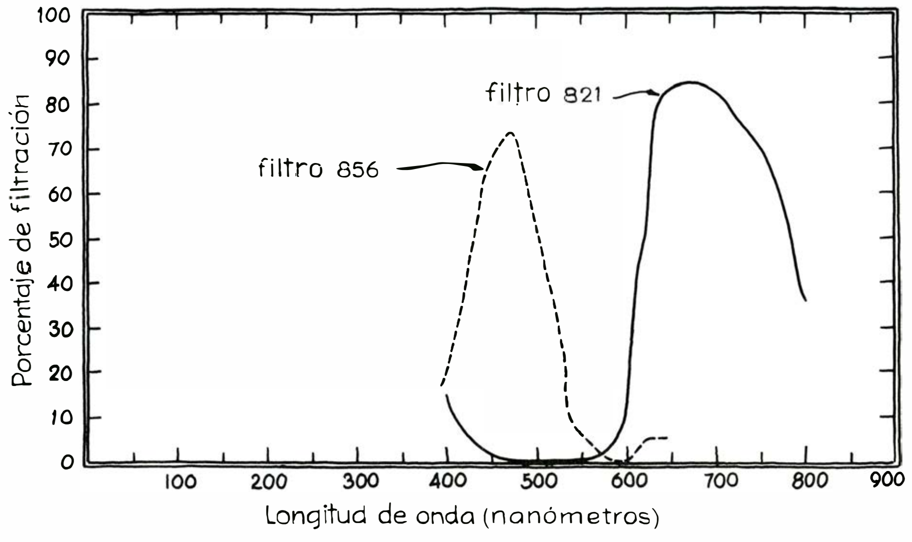


Montaje de Szabo para proyectar las diapositivas

Al mirar una escena de varios colores, su percepción resulta de dos procesos principales. Primero, el sistema visual compara las energías que le llegan en las tres ondas de máxima absorción (los receptores de longitudes de onda corta, media y larga) y asigna de acuerdo con ello una terna de números al flujo energético recibido de una determinada zona. El color no se ha determinado aún. Luego, el sistema visual compara esa zona con otras de la escena, asignando a cada una su propia terna de números. Por último, el sistema puede empezar a asignar color después de comparar las ternas de un número amplio de zonas. Por ejemplo, si una terna de una zona concreta tiene un valor alto en las longitudes de onda largas, algo más bajo en las longitudes de onda medias y aún más bajo en las cortas, se le asignará un color rojizo.

En el primero de los experimentos descritos, Szabo fotografió la tabla de colores a través de un filtro rojo; repitió luego con otro azul. El registro largo mostraba sombras blancas y negras de acuerdo con la reflexión de los distintos cuadros en las longitudes de onda rojas. Las zonas rojas aparecían claras en la filmina, las de color naranja algo menos claras y las azules relativamente oscuras. En el registro corto, la situación se invertía: las zonas azules aparecían claras, las naranjas más oscuras y las rojas eran las más oscuras de todas. Estas sombras relativas codifican la información de color que puede registrarse e interpretarse por un observador.

Cuando se proyectaba el registro largo con el mismo filtro rojo, la escena resultante era la misma (dentro de los límites de precisión de la película y la blancura de la lámpara de proyección) que la transmitida a través del filtro en la fotografía. Las zonas azules aparecían



Banda de paso a través de dos filtros

oscuras; las de color naranja, menos oscuras, y las rojas, claras. El color existente en las zonas claras era, por supuesto, sólo rosa o rojo. Cuando se apagó este proyector y se proyectó el registro corto sin filtro delante, la imagen mostraba una gama de zonas oscuras y claras reflejando lo que había sido transmitido a través del filtro azul. Esta vez la proyección sólo tenía sombras de gris oscuro, gris claro y blanco, pero registraban el grado de azul claro procedente de cada cuadro de la carta de colores.

Cuando se conectan ambos proyectores y se superponen las imágenes, ¿por qué se ven otros colores además del rosa y el rojo? Consideremos primero un cuadro que se vea rojo; en las imágenes superpuestas, éste es claro en las longitudes de onda largas, no tan claro en las longitudes de onda medianas y relativa-

Filtro de fotografía para registro corto	Filtro de fotografía para registro largo	Filtro de proyección	Comentarios
Verde 874	Rojo 821	El mismo empleado para hacer la fotografía de registro largo	Escala gris bien. Verdes pálidos. Rojos bien. Blanco bien A la inversa, cian más fuerte, mucho más brillante
Verde 877	Rojo 821		Lo mismo de antes
Verde 871	Rojo 821		Escala gris bien; verdes desvaídos, blanco; azul verdoso pálido
Verde 874, 877	Rojo 821		Buena combinación de verdes
Verde 871, 874	Rojo 821		No tan buena como la anterior
Verde 874, 877	Rojo 821, 823		Mala
Verde 874, 877	Rojo 823		Tintes marrones, rojos malos
Verde 874, 877	Naranja 819		Naranjas brillantes
Verde 874, 877	Naranja 819		Más oscuro
Azul 851	Naranja 819		Buen cian, naranja bien
Naranja 819	Rojo 821		Pálidos
Naranja 817	Rojo 821		Aun más claros
Naranja 818	Rojo 821		Muy blancos
Naranja 818	Rojo 821		Más color
Azul 851	Rojo 821		Buen verde y rojo. Ni amarillo ni violeta.
Violeta 843	Rojo 821		Más brillante que antes
Azul 856	Rojo 821		Muy buen color, un poco amarillo.
Azul 858	Rojo 821		Muy buen rojo y azul verdoso
Azul 859	Rojo 821		Más claro que antes
Azul 863	Rojo 821		No hay amarillo, los rojos y verdes buenos, no hay violeta, un poco de azul.
Azul 866	Rojo 821		Tonos color carne y una escala de grises muy buena. Color magenta pálido, no amarillo
Azul 863, 866	Rojo 821		Normal; los colores oscuros malos
Azul 863, 866	Rojo 821		Buena escala de grises y tonos color carne
Azul 863, 866	Rojo 821		Rojos y verdes azulados, no hay azul ni verde puros ni amarillo.
Amarillo, azul 807, 861	Rojo 821		Pálidos
Azul 861, 863, 866	Rojo 821		No hay amarillo, pero buen color
Azul 858	Naranja 817		Tintes anaranjados, escala gris de tono marrón
Azul 858	Amarillo 809	Naranja 817	Lo mismo que arriba
Azul 866	Verde 874, 877	Azul 866*	No tan bueno, pero aparece el amarillo
Azul 863	Verde 874, 877	Azul 863*	Mejor amarillo, la escala de grises aparece ligeramente azulada
Azul 866	Azul 863	Azul 866*	Sin efecto
Azul 859	Azul 863	Azul 863	Muy poco color
Azul 858	Azul 863	Azul 858*	Marrón y azul verdoso, escala de grises con tono marrón
Azul 856	Rojo 821	Rojo 821	Muy bueno!
Azul 856	Rojo 821 (Exposición más larga)	Rojo 821	El color más saturado a medida que la diapositiva de registro largo es menos densa. Muy buena combinación además
Azul 856	Rojo 823	Rojo 823	Sin respuesta
Azul 856 (parada más larga)	Rojo 823	Rojo 823	Mejor que antes, pero pálido; escala de grises con tono marrón
Azul 856	Naranja 817	Naranja 817	No hay rojo, sólo amarillo y naranja. Escala de grises con tono marrón, pálida
Azul 866	Verde 874, 877	Amarillo 807	Colores pálidos
Azul 866	Verde 874, 877	Verde 874, 877	Colores verdes
Azul 866	Verde 874, 877	Azul 866	Azules y amarillo pálido

* Se usa acoplador delante del registro corto

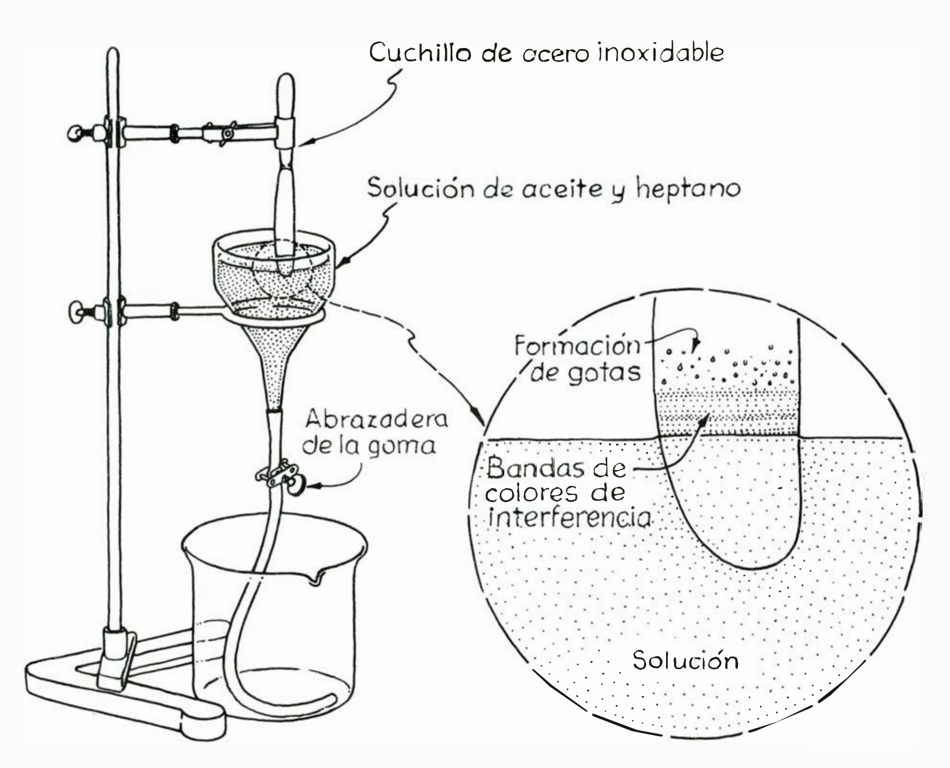
Observaciones de Szabo en una serie de experimentos

mente oscuro en las cortas, ¿por qué? Ello se explica porque el registro largo está proporcionando luz relativamente brillante en las longitudes de onda largas y el registro corto produce relativamente poca luz en las longitudes de onda cortas. En cierto sentido, el sistema visual registra los índices de claridad en cada una de las regiones de las longitudes de onda y asigna al cuadrado una terna de números de acuerdo con estos índices.

Fijémonos ahora en un recuadro que sea azul en la proyección. El registro largo está proporcionando poca luz de las longitudes de onda largas al cuadrado, mientras que el registro corto está suministrando una luz relativamente brillante. El sistema visual compara de nuevo la claridad de las longitudes de onda cortas, medias y largas y asigna una terna de números. El color azul no está asignado aún, porque el sistema visual del observador no ha entendido todavía el significado de esta terna de números con relación al color.

El proceso se repite con cada cuadrado que se ve en la proyección, al parecer de forma automática y sin la necesidad de arrastrar los ojos sobre las imágenes. A cada cuadrado se le asigna una cierta terna de números según sea la luz en las longitudes de onda cortas, medias y largas. Después se comparan estas ternas. Por ejemplo, a un cuadrado con una terna indicadora de que presenta la misma intensidad de luz en las tres regiones de longitud de onda se le asigna el color blanco, a otro cuya terna indica menos luz en las longitudes de onda medias, y aún menos en las cortas, se le asigna el rojo; a un tercer cuadrado que, en comparación con el cuadrado blanco, tenga menos luz en las longitudes de onda medias y aún menos en las largas, se le asigna el azul; a un cuadrado que sea claro en las longitudes de onda medias, un poco más oscuro en las largas y aún más oscuro en las cortas se le puede asignar el amarillo. Y así sucesivamente, hasta que se haya interpretado tanto color como sea posible de la imagen proyectada.

Supongamos ahora que la carta no tiene cuadrado blanco de referencia, ¿se destruiría por ello el proceso de asignación de colores? La respuesta es no, ya que las ternas seguirían asignadas a cada cuadrado y, por tanto, comparadas. Dado que se está enviando luz roja desde el proyector del registro largo, ésta puede actuar como punto de referencia. Por ejemplo, a una terna que sea relativamente deficiente en las longitudes de onda largas, menos pobre en las medias y muy fuerte en las cortas puede asignársele el azul en comparación con el cuadrado rojo. El experimento funciona mejor (en el sentido de que se distinguen con nitidez los colores) si los filtros usados para la fotografía tienen bandas de paso que no se solapen. De esta forma, la luz y las sombras oscuras de las filminas se distinguen mejor.



Modelo de A. A. Fote para colocar una película fina de aceite sobre una superficie de metal

La explicación de la asignación de color que he presentado aquí constituye sólo un breve resumen de un modelo más amplio desarrollado por E. Land y McCann y sus colegas. El lector interesado debería leer sus artículos para obtener así una explicación más detallada. De todas formas, dado que los experimentos de Szabo son baratos y bastante fáciles de hacer, quizás el lector se anime a repetir alguno de ellos por su cuenta.

He recibido muchas cartas sobre estereoscopios y fotografías estereoscópicas, que describí en mi artículo de febrero. Todo parece indicar que hay muchos lectores que están haciendo sus propias fotografías y filminas con esos dispositivos.

Algunos aficionados destacaban que yo mencionaba dos puntos sobre la fabricación propia de fotografías estereoscópicas. Primero, un par estereoscópico en el cual la segunda exposición se hacía girando la cámara tiene menos profundidad que uno en el que la segunda exposición se hace desplazando la cámara hacia un lado. Segundo, no se debe pensar que colocando un par de fotografías idénticas en el estereoscopio se puede lograr el mismo efecto que al colocar un par estereoscópico verdadero. Con ambos tipos de par estereoscópico el ojo puede utilizar pistas de fondo relativas al tamaño de los objetos y la geometría de las cosas, pero con fotografías idénticas es obvio que no se pueden conseguir pistas de fondo de la convergencia de los ojos.

Un motivo aparente de controversia es si desplazando la cámara entre las dos exposiciones una distancia mayor que la distancia interocular se produce una profundidad exagerada en el par estereoscópico. Un lector, Philip R. Pennington de Portland, Oregón, dice no haber visto jamás tal exageración ni siquiera en pares estereoscópicos separados por varios kilómetros. Otros lectores dicen que hacen uso regular de la distorsión de profundidad que ofrecen amplias líneas de base, sobre todo en fotografías aéreas concebidas para aumentar las variaciones de perfil de la tierra.

Unas de las muestras más bonitas de fotografías estereoscópicas que he recibido venían remitidas por James Raymond de Fairbanks, Alaska, que fotografió el Monte McKinley con una línea de base de unos 600 metros mientras volaba en un avión. He recibido también un asombroso par estereoscópico de Ralph C. Eagle, de la Facultad de Medicina de la Universidad de Pennsylvania. Eran microfotografías hechas con un microscopio electrónico que mostraban células del cáncer de una muestra de iris humano. La muestra había sido

girada varios grados entre ambas exposiciones y la profundidad se aprecia fácilmente.

Cuando hablé de los colores en las pompas de jabón en mi artículo de noviembre, describí brevemente el problema de extender una película fina en una superficie que tiene un índice de refracción más alto que la película. A. A. Fote, de la Aerospace Corporation, me ha expuesto una buena idea para conseguirlo. Con el fin de investigar el comportamiento de los aceites lubricantes bajo condiciones ambientales, preparó capas finas de aceite sobre superficies de metal. Cuando las capas son lo suficientemente finas, muestran colores de interferencia similares a los que se ven en muchas películas de jabón.

Fote mezcla una solución de aceite Apiezon con heptano de manera que el aceite sea del 20 al 40 por ciento del volumen total. Vierte la solución en un recipiente que tiene una rendija de control en el fondo; se introduce dentro del recipiente la pieza plana de metal que se desea recubrir, en posición vertical. A medida que la solución va saliendo por el fondo queda una capa fina sobre el metal. Dado que el heptano se evapora casi inmediatamente, la película sólo será de aceite. El espesor de la película de aceite puede controlarse de alguna forma cambiando la concentración del aceite en la solución y la velocidad a la que sale la solución. Las capas son más finas cuando las concentraciones son más bajas y el drenaje más lento.

Siguiendo una sugerencia de Fote sustituí el Apiezon *C* por aceite mineral ordinario, que también es soluble en heptano. Para la inmersión coloqué un embudo que drenaba a un recipiente a través de una goma, cerrada parcialmente por una abrazadera de tipo tornillo. Vertí la solución en el embudo y controlé el drenaje ajustando la abrazadera; introduje luego un trozo de metal plano y pequeño en la solución. A medida que ésta drenaba, el trozo de metal se recubría de una capa de aceite mineral. Para que la capa tenga un espesor uniforme hay que estar seguros de que la solución drena a velocidad constante. Con este fin hice mi experimento con un embudo que tenía bordes verticales por arriba y mantuve el trozo de metal en la parte superior.

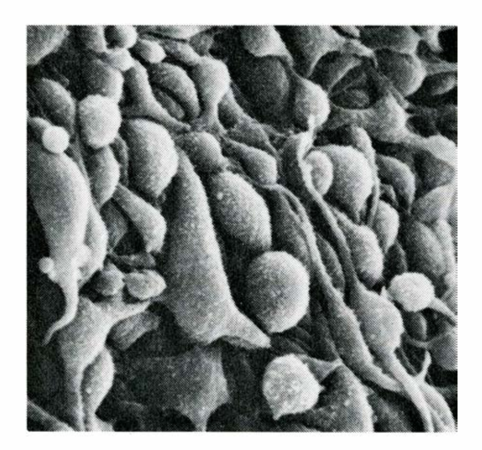
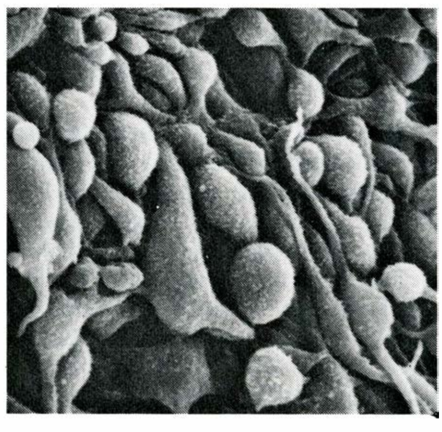
Los índices de refracción relativos de la película y el metal dominan los colores que se ven en las películas de aceite finas (si se trata de capas ultrafinas se ve una capa blanca o negra). Algunos metales tienen un índice de refracción alto (el acero 2,5 aproximadamente); otros tienen valores más bajos (el cobre 0,6 y la plata 0,2). Cuando se aplica un aceite sobre una superficie de un índice de refracción más alto, las capas ultrafinas son transparentes e incoloras. Si se aplica luz blanca en la película, ésta se refleja con el mismo color blanco al observador. Por otro lado, cuando el aceite tiene el índice de refracción mayor, la película ultrafina es oscura.

Las películas algo más espesas producen bandas de interferencia y gran variedad de colores, que probablemente el lector ha observado alguna vez en baterías de cocina metálicas. Cuando se oxida la superficie metálica de ollas y sartenes, la delgada capa de óxido actúa como una película fina y contribuye a la formación de los colores creados por la interferencia de luz. Los primeros metales que introduje en mi baño de aceite no produjeron interferencia de colores, aunque sí pude ver claramente las películas de aceite. Fote justifica este hecho diciendo que quizá mis metales no estaban suficientemente pulidos como para hacer posibles las interferencias. Existen al menos tres factores implicados en ello. Primero, las capas más finas (lo suficientemente finas para que existan interferencias perceptibles) son más probables en una superficie de metal muy lisa. Segundo, una superficie de metal muy áspera destruye la coherencia de las ondas de luz que es necesaria para la interferencia. Tercero, con una superficie áspera el espesor de la capa variará tan rápidamente de punto a punto que los colores de interferencia se mezclarán, apreciándose un color global blanco.

Para entender el segundo factor hay que seguir la trayectoria de un rayo de luz en la película. Un rayo incidente en una película se divide en dos rayos: uno que se refleja inmediatamente y otro que pasa a través de la película y se refleja por detrás de ésta. Si la superficie posterior, próxima al metal, es áspera, el rayo no se reflejará hacia fuera en la misma dirección que el rayo reflejado de la superficie anterior. Las ondas de los rayos no pueden, por tanto, interferir. Si se quiere que una superficie de metal muestre colores de interferencia en una película de aceite que la recubre, la superficie debe ser tan pulida que pueda verse la propia imagen reflejada en ella. Conseguí colores claros y muy bonitos con un cuchillo de cocina de acero inoxidable. Los colores incluían el rosa, azul y verde con algunas zonas de amarillo y marrón. Cuanto más fina era la película, más amplias se mostraban estas bandas, como ocurre con las películas de jabón.

Parece más difícil predecir los colores con estas películas de aceite que con las de jabón, porque la intensidad de la reflexión de luz de las superficies de metal varía considerablemente con la longitud de onda y con el ángulo de reflexión. Para un metal muy pulido, la reflexión de la superficie posterior de la película es mucho más fuerte que la reflexión de la frontal, de forma que los colores de interferencia se pierden en el reflejo blanco de la reflexión posterior. El reflejo parece ser muy fuerte cuando la luz incidente es aproximadamente perpendicular a la película.

Cuando sumergí por primera vez un trozo de metal en el baño de aceite (antes de empezar el drenaje) una capa subió unos 0,5 centímetros por encima de la superficie del metal. Después de un rato, la película formó "lágrimas" que descendían por el metal. En mi cuchillo de cocina pude ver colores de interferencia en la película excepto donde descendían las gotas más gruesas. La película ascendente y la formación de gotas han sido observadas en otras circunstancias desde 1850, más o menos, cuando se descubrió el efecto en vasos de vino. El ascenso suele atribuirse a que el fluido es impulsado hacia arriba por tensiones superficiales desiguales en la película vertical, un efecto bautizado con el nombre de Carlo Marangoni.

Par de microfotografías estereoscópicas de células del cáncer hechas por Ralph C. Eagle

Libros

Aportación española a la genética, un nuevo enfoque de la paleontología e introducción a la radiología

A. Prevosti, J. de Porta y J. Lafuente

GENETICA MICROBIANA. Dirección y coordinación de E. Cerdá Olmedo. Editorial Alhambra, Madrid, 1977. X + 311 páginas. Recopila trabajos presentados en un simposio sobre genética microbiana celebrado en Granada en octubre de 1973, puestos al día antes de su publicación en 1977, como lo demuestra la bibliografía que se cita. Estos trabajos, de autores españoles salvo la colaboración de Delbruck e Ingraham en dos de ellos, se presentan como capítulos y son 13 en total. La mayoría de ellos aborda un tema general, de forma sintética, e incluso los que describen y presentan procesos en un organismo o un grupo concreto tratan de poner de manifiesto su significado general.

Salas, Carrascosa, Jiménez, Camacho y Moreno, todos ellos del Departamento de Biología Molecular del Instituto Gregorio Marañón, del CSIC de Madrid, en el primer capítulo examinan la "Regulación de la expresión del ADN de virus bacterianos". Inician el trabajo con una introducción en la que ofrecen una síntesis del mecanismo de la transcripción en las bacterias, principalmente en *E. coli* y *B. subtilis,* atendiendo especialmente a las características de la ARN polimerasa que dirige dicho proceso en estas bacterias. A continuación realizan un estudio analítico de la regulación de la transcripción en los bacteriófagos T7, λ, T4 y Φ 29, extendiéndose especialmente sobre este último.

Blanco, del Laboratorio de Genética Microbiana del Instituto de Investigaciones Citológicas de la Caja de Ahorros y Monte de Piedad de Valencia, es el autor del capítulo 6, "Base genética de la lisogenia", y estudia este fenómeno en el fago λ. Al final del trabajo discute el significado que puede tener la lisogenia para la transformación de células de mamíferos por los virus productores de cáncer.

También se ocupan de un grupo taxonómico concreto Santandreu, Gascón, Gacto y Villanueva, del Departamento de Microbiología del CSIC y de la Facultad de Ciencias de la Universidad de Salamanca, que en el capítulo 10 analizan la "Regulación de la síntesis de la pared celular y enzimas exocelulares en levadura". Es un estudio principalmente bioquímico, estructural y dinámico de los polímeros que forman la membrana de las levaduras. Se consideran asimismo aspectos citológicos y del control genético de los mecanismos bioquímicos tratados.

Delbruck y Cerdá Olmedo (el primero del Instituto de Tecnología de California y el segundo del Departamento de Genética de la Facultad de Ciencias de la Universidad de Sevilla) estudian "El comportamiento de *Phycomyces*" en el capítulo 13. Describen las respuestas del esporangióforo del hongo a los estímulos externos, especialmente luminosos, y, mediante el análisis de los efectos de mutantes, hacen una disección de algunos aspectos de dichas respuestas.

Entre los capítulos que tratan temas más generales, algunos se ocupan de cuestiones esencialmente genéticas. Así, en el capítulo 2, Ramírez y del Campo, del Instituto de Biología Molecular, del CSIC (Madrid), desarrollan el tema "Regulación de la expresión génica en bacterias por AMP cíclico". Se ocupan, principalmente, de la regulación del nivel de esta substancia en la célula bacteriana y del papel que tiene en el control de la transcripción del operón *lac* en *Eschenchia coli.*

García Ballesta y Vázquez, del Instituto de Biología Celular del CSIC, Madrid, estudian la "Función y estructura del ribosoma". Toman como modelo los ribosomas de *E. coli.* Heredia, del Instituto de Enzimología del CSIC y de la Facultad de Medicina de la Universidad Autónoma de Madrid, trata de la "Iniciación de la biosíntesis de las proteínas en las bacterias". Describe los distintos elementos que realizan esta función, ARNt iniciador, ribosomas, factores de iniciación, señales del ARNm y el modo como interactúan.

La "Transformación en los procariotas" es el tema desarrollado por Portolés, López, Pérez Masiá y Espinosa, de la Sección de Biología de las Infecciones Bacterianas, Instituto Jaime Ferrán de Microbiología del CSIC, Madrid. Realizan un tratamiento muy general y amplio del tema, analizando las modalidades del fenómeno, las condiciones necesarias para que se produzca, el proceso según el cual tiene lugar y su significado para el desarrollo de la ingeniería genética.

Cánovas, Gregorio y Llamas, del Instituto de Biología celular del CSIC, estudian la "División celular en procariontes", es decir, cómo se separan los nucleoides, se forma el tabique de separación entre las células hijas y cómo éstas se separan. Además, investigan la regulación general de la división celular en las bacterias.

Asensio y del Valle, del Instituto de Enzimología del CSIC, y de la Facultad de Medicina de la Universidad Autónoma de Madrid, bajo el título "La evolución bioquímica en microorganismos", presentan una síntesis de la evolución a nivel molecular, tanto en lo que se refiere a las características de dicha evolución y de los mecanismos generales que la determinan, como a la utilización de los datos moleculares para establecer filogenias.

El capítulo dedicado al "Transporte de cationes" viene firmado por Rodríguez Navarro, del Departamento de Bioquímica del Instituto Nacional de Investigaciones Agronómicas. Estudia el transporte de cationes en microorganismos, principalmente con el objeto de explicar cómo el análisis de su base genética puede ser un medio importante para su estudio. Por ello, sólo trata ligeramente u omite otros aspectos del tema. Después de un análisis general de los mecanismos de transporte pasa al estudio de las mutaciones que afectan a los mismos.

Los capítulos séptimo y octavo son básicamente bioquímicos. El séptimo, "Reducción asimilativa de nitrato: estructura y regulación del nitrato y nitrito reductasa", por López Barea, del Departamento de Bioquímica de la Facultad de Ciencias de la Universidad de Sevilla y del CSIC, analiza el proceso en los hongos *Neurospora* y *Aspergillus.* El análisis consiste en una disección del proceso basado en el estudio de los efectos que tienen sobre el mismo diversos mutantes. El octavo, "Metabolismo de las pirimidinas", por Guerrero e Ingraham, del Departamento de Biología de la Facultad de Ciencias de la Universidad Autónoma de Barcelona y el Departamento de Bacteriología de la Universidad de California en Davis, respectivamente, constituye un completo estudio del estado del tema, principalmente en *E. coli* y *S. typhimurium* y con frecuentes referencias a otros géneros de bacterias.

En conjunto se trata de una obra útil para los interesados en la genética de microorganismos y molecular. También es útil para estudiantes de genética, microbiología y bioquímica que deseen ampliar conocimientos, pues los capítulos suelen comenzar con una introducción que hace asequible el desarrollo ulterior del tema. (A. P.).

PRINCIPIOS DE PALEONTOLOGIA, por David M. Raup y Steven M. Stanley. Traducción española de Salvador Reguant. Ariel S.A. 1.ª edición. Barcelona 1978, 456 págs. Casi sin excepción los libros y tratados de paleontología se ocupan, de una manera más o menos amplia, de la morfología y sistemática de los diferentes grupos fósiles. Las relaciones de los fósiles con su entorno geológico, el análisis de los procesos que han conducido a la formación de los conjuntos fósiles, la morfología funcional y en general todos aquellos aspectos conceptuales que constituyen el cuerpo de la paleontología fundamental o básica, han sido de manera sistemática obviados o bien tratados de una manera muy superficial. El libro de Raup y Stanley se aparta de un enfoque clásico y bien puede considerarse a estos autores como los pioneros en la sistematización de la paleontología moderna. Los autores sintetizan varios de los tópicos, que exponen en su obra, a partir de los trabajos realizados por diversos especialistas y aportando, en muchos otros casos, datos de sus propias investigaciones. Uno de los aspectos más positivos de la obra es la presentación de los diferentes temas. Varios son los ejemplos de lo que corresponde a un tratamiento superfluo. Es decir, la utilización de métodos laboriosos que no conducen a ninguna resolución y cuyo resultado puede predecirse de antemano. El capítulo de *la población como un todo,* entre otros, es muy ilustrativo en este sentido. Para mí, este tipo de enfoque y discusión de una cuestión determinada tiene la virtud de introducir en el lector un espíritu crítico del planteamiento del método que se va a seguir en la investigación. Permite de esta manera que el futuro investigador se vaya familiarizando con la no siempre fácil labor de saber adaptar y escoger, en cada caso, aquellos caracteres más apropiados en relación con el fin que se persigue. De lo contrario se cae en la rutina de la aplicación de métodos, de manera indiscriminada, que nada tienen que ver con el verdadero sentido de la investigación. Aunque a primera vista todo esto parece que no guarda relación directa con el aspecto meramente sistemático de la paleontología, es de vital importancia a la hora de realizar la descripción de una especie. Por desgracia es frecuente observar en la descripción de muchos taxones no sólo imprecisión en los caracteres, sino también, y lo que es más importante aún, la presencia de muchos datos superfluos, mientras que aquellos datos verdaderamente significativos quedan totalmente relegados o están ausentes. Bien entendido que, en el rigor científico, una cosa es la descripción de un taxón y otra el comentario o la relación que tiene un taxón con otros próximos.

Si tenemos presente que los conjuntos fósiles que se encuentran en un estrato son el resultado de una interacción de factores biológicos y geológicos, corresponde al investigador realizar la discriminación de estos factores y valorarlos concienzudamente antes de su interpretación. No pocos de los errores que se imputan a la paleontología son consecuencia de una interpretación precipitada por desconocimiento del significado y de la influencia que tienen los procesos geológicos en los conjuntos fósiles.

Es probable que algunos lectores encuentren en este planteamiento de los temas un inconveniente más que una ventaja: precisamente por la falta de dogmatismo; pero, insisto una vez más, ello obliga al lector a plantearse las dudas y a no aplicar una interpretación rígida. Al fin y al cabo ni los procesos biológicos ni los geológicos responden a un esquema rígido.

La observación del tipo de distribución tamaño/frecuencia en un conjunto fósil permite darse cuenta de hasta qué punto están conjugados los procesos biológicos y geológicos. Una atenta lectura de los capítulos 3 y 4 harán meditar a muchos investigadores acerca del valor real que tiene esta paleontología básica o fundamental, tantas veces desdeñada, a la hora de interpretar los resultados.

La obra de Raup y Stanley está concebida en dos partes. La primera en relación con el tratamiento del concepto de fósil bajo sus aspectos más significativos, tanto en su contexto geológico como biológico. Esta primera parte comprende siete capítulos. En ellos se trata, entre otros, la formación y destrucción de los fósiles, la variación ontogenética y el crecimiento, el concepto de especie y los tipos de especiación. Singular importancia tiene el tratamiento que los autores dan a la dinámica de las poblaciones y a la información, aunque breve, sobre el Código Internacional de Nomenclatura Zoológica, así como la disposición y tratamiento de las sinonimias. Se incluye también una orientación acerca de la taxonomía numérica y los criterios para la definición de las categorías taxonómicas de mayor rango. En la segunda parte se describe la utilización de los datos paleontológicos. Además de los conceptos que se emplean en la bioestratigrafía, destacan los capítulos sobre la adaptación y morfología funcional, campo en el que los autores tienen varias publicaciones, paleoecología y evolución. En el capítulo de paleoecología, entre los tópicos más significativos que tienen mayor trascendencia para la interpretación de medios sedimentarios, figura la relación entre la composición de la fauna y los diferentes tipos de sustratos. Por último, los procesos diagenéticos se incluyen en el apartado de la geoquímica.

La obra de Raup y Stanley está redactada en un estilo sencillo, conciso y es de lectura amena. La parte gráfica, muy bien ilustrada, con ejemplos muy bien escogidos tanto en el campo de los invertebrados como en el de los vertebrados. Al final de la obra se incluye una amplia bibliografía que está completada por una bibliografía más específica al término de cada uno de los diferentes capítulos.

Es lógico que, dado el amplio campo que abarca la paleontología bajo esta óptica, no se puedan tratar todos los aspectos, ni ninguno de ellos, de forma exhaustiva. Dicho sea de paso que no es la intención de los autores realizar una obra de gran envergadura, como manifiestan claramente en el prólogo, pero sí quizá se notan a faltar, en mi opinión, determinados aspectos de la actividad biológica. Aspectos que se encuentran bien representados en los conjuntos fósiles: las perforaciones y las pistas. De ambas manifestaciones se encuentran datos en diferentes capítulos; sin embargo una sistematización en un capítulo, por pequeño que fuera, pero independiente, hubiera dado el verdadero valor que tienen.

Aunque el libro está dedicado a los conocimientos fundamentales para un nivel de licenciatura, resulta muy útil también para los paleontólogos y geólogos en general. Añadiría que de forma muy especial para aquellos que aún siguen aferrados a un concepto clásico, que ven como objetivo exclusivo de la paleontología la identificación de los fósiles para las dataciones cronoestratigráficas.

Como señalan los autores en su prólogo, el libro no pretende cubrir un curso completo de paleontología. La intención de plantear una visión amplia, con una nueva metodología que despierte la imaginación del futuro investigador, está plenamente conseguida. Al mismo tiempo abre un amplio abanico de posibilidades en este campo.

La versión española es correcta. Quizá se pueden hacer algunas observaciones a determinadas expresiones como la de "peso de playa", como equivalente de "sand dollar", pero en general está bien cuidada, lo mismo que la impresión de la obra. En la bibliografía se han introducido algunas citas de carácter general o de carácter específico de algún tema, correspondientes a autores de habla hispánica.

Por último hay que lamentar el retraso en la aparición de esta versión española que viene a coincidir con la aparición de la segunda edición, ampliada en muchos aspectos, que acaban de realizar los mismos autores. Con todo, la traducción española viene a llenar un vacío muy importante en este campo (J. P.)

RADIOLOGIA GENERAL. COMPENDIO ANALITICO-SEMIOLOGICO, por Isadore Meschan. Primera edición española de la obra "Synopsis of analysis of Roentgen signus in General Radiology". Editorial A C. Madrid. 1978, 663 págs. La obra original en tres volúmenes, cuya última edición de W. B. Sanders Company ha sido la de 1976, es un clásico en la literatura radiológica americana. El profesor Isadore Meschan, Director del Departamento de Radiología de la Bowman Gray School of Medicine, de Carolina del Norte, mantiene en su obra original una unidad que, lamentablemente, falta en este compendio. En palabras de su autor, el objetivo del libro es "la catalogación e ilustración de los principales signos radiológicos y su clarificación y correlación con el diagnóstico y la práctica clínica".

Considerada desde el punto de vista didáctico, se llega a la conclusión de que la obra es un compendio de conocimientos básicos, útiles para el estudiante que se inicia en el radiodiagnóstico. Los esquemas que ilustran el texto, perfectos en su precisión, suponen un resumen de la misma y, dada su planificación, cada capítulo ofrece a su término un amplio cuestionario que favorece la autoevaluación. Evitando complejidades, y manteniendo su sistemática de estudio, el libro expone los principios fundamentales del radiodiagnóstico y abarca desde la naturaleza de los Rx y la geometría de la imagen hasta el uso de los contrastes y las nociones de radioprotección.

Los primeros capítulos, dedicados a la nomenclatura radiológica, ofrecen al estudiante la posibilidad de aprender a definir con exactitud las imágenes radiográficas y a asimilar, a su vez, la terminología anatomopatológica. En los siguientes capítulos, con las limitaciones de extensión obligadas, el autor hace un recorrido de la radiología diagnóstica por aparatos; así, en el capítulo 4, como introducción al sistema músculo-esquelético, estudia la fisiopatología y bioquímica del hueso, su método de análisis radiológico y el desarrollo y maduración del esqueleto.

El examen de las fracturas y luxaciones de las extremidades, tema del capítulo 5, con la aportación, ya mencionada, de unos buenos y clarificadores esquemas se sigue de un estudio sencillo de la evolución, complicaciones y métodos de tratamiento de las fracturas, desde el punto de vista radiográfico. La parte referida a anomalías congénitas y hereditarias, siendo algo más extensa, se mantiene reducida, ya que no compendia las múltiples clasificaciones propias de otros tratados, y tan sólo ejemplariza algunas anomalías frecuentes.

Es, por otro lado, muy apropiada la comparación del contorno interior de la pelvis con diversos tipos de copas: copa de coñac para la pelvis normal, de champán para la acondroplasia y de vino en la enfermedad de Morquio. No obstante, el estudio de la mano, tan fundamental para la tipificación de anomalías congénitas, ha sido olvidado. Continúan los capítulos 8, 9 y 10 con un estudio semiológico de la patología ósea, en la que se subraya la importancia de la valoración del signo para llegar al diagnóstico. A partir de ahí, relaciona una clasificación de las enfermedades según su radiotransparencia o radioopacidad, difusa o circunscrita, generalizada o con afectación unitaria, con o sin esclerosis marginal de la lesión, etcétera. En la clasificación y estudio sistemático de las articulaciones se incluyen dibujos de gran sentido patogénico.

Las proyecciones básicas y la anatomía radiológica del cráneo sirven de introducción al capítulo de signos radiológicos de las anomalías craneales. El énfasis que se hace sobre el escaso valor de las radiografías simples en el diagnóstico de las fracturas de cráneo es de gran interés, por la interpretación de que, en general, no hay correlación entre la posible fractura y los signos clínicos, no influyendo en el tratamiento ni en el pronóstico, salvo en los casos de fractura deprimida o combinada. Puntualiza también que la realización de radiografías al niño asintomático, con trauma craneal, eleva innecesariamente la dosis de radiación a la población infantil. A continuación se compendian brevemente algunas regiones específicas del cráneo: silla turca, senos paranasales, mastoides, etcétera. Pone fin al estudio del sistema óseo el capítulo 14, dedicado a la columna vertebral.

El análisis radiológico del tórax, incluyendo el diafragma, pleura, caja torácica, vías aéreas superiores, parénquima pulmonar y corazón es el tema que conforma los capítulos 15 al 23. Junto con los ya conocidos gráficos de la distribución lobar y segmentaria, se incluyen los nuevos conceptos sobre lobulillo primario y secundario de tanta importancia en el diagnóstico diferencial de las enfermedades pulmonares generalizadas. La patología pulmonar es estudiada semiológicamente en distintos capítulos, partiendo de las características morfológicas de las lesiones básicas. Faltan, no obstante, en el capítulo dedicado al mediastino, los modernos conceptos de las líneas "para", y la clasificación radiológica de los compartimentos. En los dedicados a corazón (cardiopatías congénitas y adquiridas), la gran claridad de los dibujos y los numerosos "gamuts" los vuelven altamente comprensibles.

Proporcionalmente a otras regiones, al estudio del abdomen se le ha dado mayor amplitud que a otras regiones, y el dedicado al abdomen sin contraste es muy completo, pese a estar ausentes algunos conceptos de la radiología moderna. En el capítulo 25 encontramos la metodología radiológica del riñón y las cápsulas suprarrenales, correctamente expuestas. Sin embargo, hemos de considerar insuficiente el dedicado al estudio del sistema biliar.

Sucesivamente se va pasando al estudio semiológico del tracto digestivo, y son aprovechados algunos esquemas del libro de Margulis, para exponer la patología esofágica y gástrica; ocasionalmente, se introducen en el texto unos recuadros en fondo gris con ideas actualizadas y cuadros resumen.

En la parte dedicada al intestino delgado el autor incluye ejemplos de enfermedades específicas recogidas de textos de otros autores importantes. Completa esta obra el diagnóstico radiológico en obstetricia y ginecología, con las características generales de sencillez y esquematismo. Así pues el libro, dedicado al estudiante, cumple su misión de iniciar en la radiología diagnóstica y facilitar la consulta rápida para recordar conceptos de manejo diario. Sin olvidar que la radiología moderna, con la introducción de técnicas nuevas (ecografía y tomodensitometría), ha dado un gran avance que no está recogido en este libro. El volumen, bien editado por A C, se presenta en papel mate, lo que supone una pérdida en la reproducción de radiografías, y disminuye, en general, la calidad iconográfica. Sin embargo, su llegada, dada la poca literatura radiológica que actualmente disponemos en nuestro país en nuestro propio idioma, es muy de agradecer. (J. L.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

LA FERTILIZACION HUMANA EXTERNA

HAMILTON, BOYN AND MOSSMAN'S HUMAN EMBRYOLOGY: PRENATAL DEVELOPMENT OF FORM AND FUNCTION. 4.ª edición por W. J. Hamilton y H. W. Mossman. W. Heffer & Sons Ltd. 1972.

FERTILIZATION OF HUMAN EGGS IN VITRO: MORALS, ETHICS AND THE LAW. R. G. Edwards en *The Quarterly Review of Biology,* vol. 49, n.º 1, págs. 3-26; marzo, 1974.

RESEARCH ON THE FETUS. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Department of Health, Education, and Welfare Publication n.º (OS) 76-128, 1975.

EL FACTOR DE CRECIMIENTO NERVIOSO

THE NERVE GROWTH FACTOR: ITS MODE OF ACTION ON SENSORY AND SYMPATHETIC NERVE CELLS. Rita Levi-Montalcini en *The Harvey Lectures,* Series 60, págs. 217-259; 1966.

NERVE GROWTH FACTOR FROM MOUSE SUBMAXILLARY GLAND: AMINO ACID SEQUENCE. Ruth Hogue Angeletti y Ralph A. Bradshaw en *Proceedings of the National Academy of Sciences of the United States of America,* vol. 68, n.º 10, págs. 2417-2420; octubre, 1971.

THE RETROGRADE AXONAL TRANSPORT OF NERVE GROWTH FACTOR. I. A. Hendry, K. Stöckel, H. Thoenen y L. L. Iversen en *Brain Research,* vol. 68 n.º 1, págs. 103-121; 15 de marzo de 1974.

NERVE GROWTH FACTOR. William C. Mobley, Alfred C. Server, Douglas N. Ishi, Richard J. Riopelle y Eric M. Shooter en *The New England Journal of Medicine,* vol. 297, n.º 20, págs. 1096-1104, 17 de noviembre de 1977; vol. 297, n.º 21, págs. 1149-1158, 24 de noviembre de 1977; vol. 297, n.º 22, págs. 1211-1218, 1 de diciembre de 1977.

NERVE GROWTH FACTOR POTENTIATES ACTOMYOSIN ADENOSINETRIPHOSPHATASE. Pietro Calissano, Gianni Monaco, Loriana Castellani, Delio Mercanti y Andrea Levi en *Proceedings of the National Academy of Sciences of the United States of America,* vol. 75, n.º 5, págs. 2210-2214; mayo, 1978.

CONTROL AUTOMATICO POR INTELIGENCIA DISTRIBUIDA

DIRECT DIGITAL PROCESS CONTROL: PRACTICE AND ALGORITHMS FOR MICROPROCESSOR APPLICATION. D. M. Auslander, Y. Takahashi y Y. Tomiyuka en *Proceedings of the IEEE,* vol. 66, n.º 2, págs. 199-208; febrero, 1978.

SPECIAL ISSUE ON LARGE-SCALE SYSTEMS AND DECENTRALIZED CONTROL. *IEEE Transactions on Automatic Control,* vol. AC-23, n.º 2; abril, 1978.

TEORIA DE NUDOS

KNOTS AND LINKS. Dale Rolfsen. Publish or Perish, Inc., Berkeley, Calif., 1976.

INTRODUCTION TO KNOT THEORY. Richard H. Crowell y Ralph H. Fox. Springer-Verlag, 1977.

UNA MINA DE SILEX NEOLITICA

LA STATION NÉOLITHIQUE DE SAINTE-GERTRUDE. Joseph Hamal-Nandrin y Jean Servais en *Revue Anthropologique,* vol. 33, págs. 345-492; 1923.

DER ÄLTESTE BERGBAU IN EUROPA. Martin Jahn. Akademie-Verlag, Band 52, Heft 2, 1960.

5000 JAHRE FEUERSTEINBERGBAU IN DEN NIEDERLANDEN. F. H. G. Engelen en *Der Anschnitt,* vol. 21, n.º 2, 1969.

NEUTRONES ULTRAFRIOS

ELECTRIC DIPOLE MOMENT OF THE NEUTRON. R. Golub y J. M. Pendlebury en *Contemporary Physics,* vol. 13, n.º 6, págs. 519-558; noviembre, 1972.

ULTRACOLD NEUTRONS. V. I. Luschikov en *Physics Today,* vol. 30, n.º 6, págs. 42-51; junio, 1977.

ULTRA-COLD NEUTRONS. R. Golub y J. M. Pendlebury en *Reports on Progress in Physics.* Institute of Physics, marzo, 1978.

LA HISTORIA DEL ATLANTICO

AN ANALYSIS OF THE VARIATION OF OCEAN FLOOR BATHYMETRY AND HEAT FLOW WITH AGE. Barry Parsons y John G. Sclater en *Journal of Geophysical Research,* vol. 82, n.º 5, págs. 803-827; 10 de febrero de 1977.

THE PALEOBATHYMETRY OF THE ATLANTIC OCEAN FROM THE JURASSIC TO THE PRESENT. John G. Sclater, Steven Hellinger y Christopher Tapscott en *The Journal of Geology,* vol. 85, n.º 5, págs. 509-552; septiembre, 1977.

LAS MICROCINAS

A NEW FAMILY OF LOW MOLECULAR WEIGHT ANTIBIOTICS FROM ENTEROBACTERIA. C. Asensio, J. C. Pérez-Díaz, M. C. Martínez y F. Baquero, en *Biochem. Biophys. Res. Commun.,* vol. 69, págs. 7-14, 1976.

ECOLOGIA MOLECULAR. C. Asensio, en *Avances de la Bioquímica,* editado por L. Cornudella y col., Salvat, Barcelona, 1977.

MICROCIN PLASMIDS: A GROUP OF EXTRACHROMOSOMAL ELEMENTS CODIFYING FOR LOW MOLECULAR ANTIBIOTICS ESCHERICHIA COLI. F. Baquero, D. Bouanchaud, M. C. Martínez-Pérez y C. Fernández, en *J. of Bacteriol.* vol. 135, págs. 342-437, 1978.

EXTRACHROMOSOMALLY DETERMINED ANTIBIOTIC PRODUCTION. D. A. Hopwood, en *Ann. Rev. Microbiol.,* vol. 32, págs. 373-392, 1978.

JUEGOS MATEMATICOS

ON COLORING THE $N \times N$ CHESSBOARD. M. R. Iyer y V. V. Menon en *The American Mathematical Monthly,* vol. 73, n.º 7, págs. 721-725; septiembre, 1966.

INDEPENDENT PERMUTATIONS, AS RELATED TO A PROBLEM OF MOSER AND A THEOREM OF PÓLYA. Ashok K. Chandra en *Journal of Combinatorial Theory, Series A.* vol. 16, n.º 1, págs. 111-120; enero, 1974.

TALLER Y LABORATORIO

INTERFERENCE COLOR. Hiroshi Kubota en *Progress in Optics: vol. 1,* dirigido por E. Wolf. North-Holland Publishing Company, 1961.

LIGHTNESS AND RETINEX THEORY. Edwin H. Land y John J. McCann en *Journal of the Optical Society of America,* vol. 61, n.º 1, págs. 1-11; enero, 1971.

Seguiremos explorando los campos del conocimiento



LOS EFECTOS INMEDIATOS Y RETARDADOS DE LA GUERRA NUCLEAR, por Kevin N. Lewis
La importancia y potencia de las fuerzas estratégicas de los Estados Unidos se fundamentan en los efectos inmediatos producidos por las armas nucleares. Igualmente importantes resultan los efectos retardados de las mismas y todo ello obliga a garantizar el que esas fuerzas tengan que mantener su papel de elemento disuasorio al más alto nivel.

EL VIRUS DE EPSTEIN-BARR, por Werner Henle, Gertrude Henle y Evelyne T. Lennette
Es posiblemente el más común de todos los virus que infectan a los seres humanos. Provocan la mononucleosis infecciosa, y se poseen sólidas pruebas que lo relacionan con dos tipos de cáncer humano.

EL PARSEC CENTRAL DE LA GALAXIA, por Thomas R, Geballe
Las observaciones de radio y en el infrarrojo indican que el centro de nuestra galaxia aloja un objeto ultracompacto, posiblemente un enorme agujero negro, inmerso en una densa masa de estrellas, gas y polvo en rotación.

LOS ORIGENES DEL PRIMER AVION MOTOPROPULSADO CON PILOTO A BORDO, F. E. C. Culick
El avión de los hermanos Wright de 1903 no fue sólo un esfuerzo afortunado de dos mecánicos de bicicletas de Dayton, sino el resultado de un programa intensivo de investigación, ingeniería y experimentación.

COMPARTIMENTOS EN EL DESARROLLO DE LOS ANIMALES, por Antonio García-Bellido, Peter A. Lawrence y Ginés Morata
La mosca del vinagre, y es posible que también otros animales, está construida a base de unas unidades llamadas compartimentos.

EL MODELO BOLSA DEL CONFINAMIENTO DE LOS QUARKS, por Kenneth A. Johnson
Los quarks parecen ser reales, a pesar de que todavía no se han visto aislados. Una hipótesis sobre su "ocultación" es la que los confina en bolsas.

LOS ANIMALES DE BURGESS SHALE, por Simon Conway Morris y H. B. Whittington
Los fósiles de una formación rocosa en el Canadá occidental son una rica muestra de una comunidad animal del Cámbrico medio.

FREGE, por Javier de Lorenzo
El año 1879 constituye una fecha clave en la historia de la lógica matemática: la de su nacimiento. Frege publicó hace ahora exactamente un siglo su Ideografía.

